

The
University
Of
Sheffield.

Access
To
Thesis.

This thesis is protected by the Copyright, Designs and Patents Act 1988. No reproduction is permitted without consent of the author. It is also protected by the Creative Commons Licence allowing Attributions-Non-commercial-No derivatives.

- A bound copy of every thesis which is accepted as worthy for a higher degree, must be deposited in the University of Sheffield Library, where it will be made available for borrowing or consultation in accordance with University Regulations.
- All students registering from 2008–09 onwards are also required to submit an electronic copy of their final, approved thesis. Students who registered prior to 2008–09 may also submit electronically, but this is not required.

Author: Andrew Booth

Dept: SCHARR

Thesis Title: Acknowledging a "Dual Heritage" for Qualitative E

Registration No: 110279107

For completion by all students:

Submit in print form only (for deposit in the University Library):

☐

Submit in print form and also upload to the *White Rose eTheses Online* server:

In full

☒

Edited eThesis

☐

Please indicate if there are any embargo restrictions on this thesis. Please note that if no boxes are ticked, you will have consented to your thesis being made available without any restrictions.

Embargo details: (complete only if requesting an embargo to either your print and/or eThesis)

Embargo required?

Length of embargo
(in years)

Print Thesis Yes ☐ No ☒

eThesis Yes ☒ No ☐

1

Supervisor: I, the supervisor, agree to the named thesis being made available under the conditions specified above.

Name: Dr Eva Kaltenthaler

Dept: SCHARR

Signed: *Eva Kaltenthaler*

Date: 2/1/2013

Student: I, the author, agree to the named thesis being made available under the conditions specified above.

I give permission to the University of Sheffield to reproduce the print thesis in whole or in part in order to supply single copies for the purpose of research or private study for a non-commercial purpose.

I confirm that this thesis is my own work, and where materials owned by a third party have been used copyright clearance has been obtained. I am aware of the University's *Guidance on the Use of Unfair Means* (www.sheffield.ac.uk/lets/design/unfair)

I confirm that all copies of the thesis submitted to the University (including electronic copies on CD/DVD) are identical in content.

Name: Andrew Booth

Dept: SCHARR

Signed: *A. Booth*

Date: 7/1/2013

For completion by students also submitting an electronic thesis (eThesis):

I, the author, agree that the University of Sheffield's eThesis repository (currently WREO) will make my eThesis available over the internet via an entirely non-exclusive agreement and that, without changing content, WREO may convert my thesis to any medium or format for the purpose of future preservation and accessibility.

I, the author, agree that the metadata relating to the eThesis will normally appear on both the University's eThesis server and the British Library's ETHOS service, even if the thesis is subject to an embargo. I agree that a copy of the eThesis may be supplied to the British Library.

I confirm that the upload is identical to the final, examined and awarded version of the thesis as submitted in print to the University for deposit in the Library (unless edited as indicated above).

Name: Andrew Booth

Dept: SCHARR

Signed: *A. Booth*

Date: 7/1/2013

THIS SHEET MUST BE BOUND IN THE FRONT OF THE PRINTED THESIS BEFORE IT IS SUBMITTED

Acknowledging a 'Dual Heritage' for Qualitative Evidence Synthesis: Harnessing the Qualitative Research and Systematic Review Research Traditions

Andrew Booth

**Thesis submitted to the University of Sheffield for the
degree of Doctor of Philosophy by Publication**

School of Health and Related Research

June 2013

Table of Contents

Acknowledgements.....	4
An Explanation of Terminology	5
Summary	6
Abbreviations.....	7
Glossary	8
Relevant publications and declaration of contribution.....	10
Methodological papers	10
Case studies as methodology lead.....	13
Coherence of Work	14
1. Background.....	15
1.1 Qualitative Systematic Reviews: same family, different heritage?.....	17
1.2 Towards a 'Dual Heritage' model of Qualitative Evidence Synthesis	18
2. Context for the Featured Papers.....	23
3. On Searching and Sampling.....	25
3.1 Limitations of Papers M1 and M2.....	32
3.2 Summary	33
4. On Quality Assessment	34
4.1 Role of the Case Studies	38
4.2 Limitations of Paper M3	40
4.3 Summary	42
5. On Synthesis	43
5.1 Towards a pragmatic method of synthesis	44
5.2 Exploring heterogeneity	45
5.3 Role of the Case Studies	46
5.4 Limitations of Papers M4 and M5.....	49
5.5 Summary	51
6. Improving Standards of Reporting	52
6.1 Summary.....	55

A 'Dual Heritage' for QES

7. Some Reflections on the Dual Heritage model.....	57
8. Towards a Future Research Agenda.....	59
9. Conclusion.....	62
References.....	66
Appendix A – Summary of Permissions.....	82
Appendix B - Paper M1 (“Brimful of Starlite”).....	83
Appendix C - Paper M2 (Synthesizing Qualitative Research).....	84
Appendix D - Paper M3 (Should We Exclude Inadequately-Reported Studies?).....	85
Appendix E - Paper M4 (A Worked Example of “Best Fit” ...Synthesis).....	86
Appendix F - Paper M5 (Desperately Seeking Dissonance).....	87
 Figure 1 - Relationship between Included Papers.....	 12
 Table 1 - Potential Contributions to Qualitative Evidence Synthesis of the respective heritages.....	 22
Table 2 - Synthesis Methods with Appropriate Sampling Methods.....	30
Table 3 - Strategies for exploring heterogeneity within a QES (from Paper M5).....	46
Table 4 - Characteristics of the Case Studies.....	48
Table 5 - Elements of the STARLITE mnemonic (from Paper M1).....	53

A 'Dual Heritage' for QES

Acknowledgements

Firstly, I would like to thank **Dr Eva Kaltenthaler** for being an excellent and most supportive advisor throughout my PhD study. I would also like to thank my co-convenors of the Cochrane Collaboration Qualitative Research Methods Group, **Professor Jane Noyes, Dr Karin Hannes and Dr Janet Harris**, for acting as external readers and **Professor Alicia O'Cathain, Dr Christopher Carroll, Dr Emma Knowles and Dr Susan Baxter** for acting as ScHARR readers. Indeed particular thanks are due to **Christopher Carroll** with whom innumerable kitchen and corridor conversations have helped turn our opportunistic methodology research into a reality.

I have benefited greatly over the past decade from the challenging debates and discussions I have had with colleagues on various funded projects, with a particular debt to **Professor Mary Dixon-Woods and the University of Leicester Quantitative and Qualitative Review Methods Team** who helped me to embark on my journey of discovery. Thank you too to all my co-authors on the included case studies.

Thanks to my family, wife **Sandra** and children, **James, Thomas and Eloise** for tolerating my extended absence spent in the garret that is our home office.

Finally this PhD by Publications was undertaken during a period of University of Sheffield supported study leave (April 2012 – September 2012) and I would like to thank **Professors Alan Brennan, Simon Dixon and Jon Nicholl** for being instrumental in securing this protected time.

A 'Dual Heritage' for QES

An Explanation of Terminology

Although systematic reviews constitute a collective endeavour, this Thesis highlights a unique, original and personal contribution to knowledge of evidence synthesis. The following reporting conventions support this focus. The five featured Methodological papers are designated with the prefix of M e.g. **Paper M1**. Methodological issues are exemplified through five case studies labelled as **Case Study 1** etcetera. The Methodological papers are indicated in the References with the entire reference in bold accompanied by three asterisks e.g. *** **Booth A (2006) "Brimful of Starlite"** etc.

Additional relevant supporting papers that underpin the Candidate's personal contribution are indicated within the **References** by the Candidate's name being highlighted in bold e.g. "Carroll C, **Booth A**, Papaioannou D, Sutton A, & Wong R (2009a) etc". Furthermore such references are indicated within the body of the text by the addition of the Candidates's name to the First Author et al convention, with ellipsis to indicate authorial position subsequent to second author, e.g. "Dixon-Woods...Booth et al (2006)". This convention enables the reader to place the featured papers within a programme of interconnected research and yet indicates transparently the extent of engagement with published work by, and with, other researchers.

Finally, although 'qualitative systematic review' persists as the most common umbrella term for the methodology described in this Thesis it carries unwanted connotations that the included papers seek to challenge, namely that qualitative research can be mechanistically 'squeezed' within a conventional systematic review template. Unless constrained by the terminology used in source articles, this Thesis implements the decision, taken by the Author with other co-convenors of the Cochrane Collaboration Qualitative Methods Group, to privilege the term 'qualitative evidence synthesis' (QES).

A 'Dual Heritage' for QES

Summary

Qualitative evidence synthesis, an umbrella term that includes all forms of secondary qualitative synthesis and analysis including **qualitative systematic review**, has emerged from the confluence of conventional systematic review methods with methods for primary qualitative research. With such a 'mixed heritage', and the juxtaposition of quite different epistemological positions, it is inevitable that the resultant tensions have generated considerable creative energy and significant methodological frictions.

These tensions have created an environment within which I have sought to make a contribution. Working with colleagues within the School of Health and Related Research (ScHARR), University of Sheffield, and collaborators at other institutions, including fellow co-convenors of the Cochrane Collaboration Qualitative Methods Group, I have examined the state of qualitative synthesis methods (**Paper M1**). I have traced and examined the respective contributions of the two components of the mixed heritage through five methodological papers that examine the stages of the systematic review process; searching (**Paper M2**), quality assessment (**Paper M3**), framework and thematic synthesis (**Paper M4**) and exploring heterogeneity (**Paper M5**) through to consideration of reporting standards (**Papers M1-M3**). This Thesis explores these issues through five case studies (**Case Studies 1-5**) to which I have contributed as lead methodologist.

While, initially at least, the legacy of conventional systematic review methods could be seen to enjoy dominance, an emerging imperative to review systematically different types of evidence to explore different review questions, coupled with reduced time and resource envelopes within which to address time-critical questions from policy and practice, has opened up a more versatile and pragmatic toolkit. The Thesis concludes by identifying key methodological issues that require further investigation. I contend that many outstanding methodological challenges may derive their most productive insights from a more detailed consideration of corresponding solutions from primary qualitative research. The five papers in this body of work, therefore, make an original contribution to knowledge by establishing and demonstrating methodological principles by which flexible and context sensitive application of the versatile 'systematic review model' can be used to meet the pragmatic demands of health services research and technology assessment.

Word Count: 14,382 (Excluding Tables, Appendices and Supporting Material)

A 'Dual Heritage' for QES

Abbreviations

CIS	- Critical Interpretive Synthesis
COREQ	- Consolidated criteria for reporting qualitative research
ENTREQ	- ENhancing Transparency in Reporting the synthesis of Qualitative research
EQUATOR	- Enhancing the QUality And Transparency Of health Research
ESRC	- Economic and Social Research Council (UK)
GS	– Google Scholar
IPA	- Interpretative Phenomenological Analysis
NHS	- National Health Service
PICO	– Population (or Patient) Intervention Comparison Outcome
PICOC	- Population Intervention Comparison Outcome Context
PICOS	- Population Intervention Comparison Outcome Study type
PRISMA	- Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QES	– Qualitative Evidence Synthesis
SALSA	- Search Appraisal Synthesis Analysis
SPICE	- Setting Perspective (phenomenon of Interest) Comparison Evaluation
SPIDER	- Sample, Phenomenon of Interest, Design, Evaluation, Research type
SQUIRE	- Standards for Quality Improvement Reporting Excellence
SRs	- Systematic Reviews
STARD	- STAndards for the Reporting of Diagnostic accuracy studies
STARLITE	- Sampling strategy, Type of study, Approaches, Range of years, Limits, Inclusion and exclusions, Terms used, Electronic sources
WoS	- Web of Science

A 'Dual Heritage' for QES

Glossary

Aggregative review - a type of review that is concerned with assembling and pooling data (either quantitative as with meta-analysis or qualitative as with thematic synthesis). To achieve such aggregation requires that there is basic comparability between phenomena.

Best fit synthesis – a variant of **framework synthesis**, subsequently labelled by Booth (**Paper M4** - Carroll, Booth & Cooper, 2010), that combines the inherent advantages of inductive and deductive approaches to thematic analysis. It involves a contingent approach to using a model or framework until emerging themes are used to enhance or augment the original framework.

Configurative review – a type of review that seeks to interpret and understand the world by interpreting and arranging (configuring) information and by developing concepts

Critical interpretive synthesis - an approach to the synthesis of qualitative and quantitative data used in situations where theorisation of the evidence is required. Critical interpretive synthesis encourages a critique of literatures and the questioning of taken-for-granted assumptions about concepts and methods

Dual heritage – the fact of having parents from different ethnic/cultural backgrounds; used metaphorically in this Thesis to refer to qualitative evidence synthesis as the product of two research traditions, namely those of primary qualitative research and systematic review.

Framework synthesis - a process of synthesis, analogous to the use of framework analysis, but used to analyse data from multiple studies within a review.

Interpretive review - a type of review that seeks to use the process of synthesis as a means of explaining a particular phenomenon. Gough et al (2012a; 2012b) and others prefer the accuracy conveyed by “configurative review”

Meta-ethnography- the most common specific method of synthesis of qualitative research (Major & Savin-Baden, 2011; Hannes & Macaitis, 2012), originally used to synthesise ethnographies but now used to refer to synthesis of other study types, typically with the objective of theory generation.

Meta-interpretation – an approach to qualitative evidence synthesis, proposed by Weed (2005), that fuses emergent and iterative (i.e. not predefined) techniques to process interpretations, as the raw data of synthesis, with a requirement for a transparent audit trail.

Meta-narrative synthesis - a type of synthesis that seeks to explore large and heterogeneous literatures from different research traditions or disciplines by following the unfolding storyline or narrative from each.

Narrative synthesis – a systematic process of describing the shared properties of a group of studies included in a review primarily through text but augmented by tabular and graphical displays of data.

Qualitative evidence synthesis (QES) - an umbrella term, endorsed by the Cochrane Collaboration Qualitative Methods Group, increasingly used to describe a group of

A 'Dual Heritage' for QES

review types that attempt to synthesise and analyse findings from primary qualitative research studies.

Quality assessment - the systematic process of examining the internal and external validity, together with other considerations such as **dependability and neutrality**, of studies for potential inclusion in a review so as to evaluate their individual contributions to the overall 'bottom line' of that review.

Realist synthesis - a method for studying complex interventions in response to the perceived limitations of conventional systematic review methodology. It involves identification of contexts, mechanisms and outcomes for individual programmes in order to explain differences, intended or unintended, between them.

SALSA - an acronym and mnemonic devised by Booth, Papaioannou & Sutton (2011) describing the processes required, to varying degrees and intensities, in order to achieve a systematic approach to reviewing the literature according to the elements of **Search, Appraisal, Synthesis and Analysis**.

Scoping review – a type of review that has as its primary objective the identification of the size, nature and quality of research in a topic area in order to inform a subsequent review.

Sensitivity analysis - An analysis used to determine how sensitive the results of a study or systematic review are to changes in how it was done
<http://www.jr2.ox.ac.uk/bandolier/booth/glossary/sensanal.html>

Study identification - the process of searching and retrieving an initial set of documents and then deciding on their eligibility in order to arrive at a manageable set of includable studies.

Relevant publications and declaration of contribution

The Candidate is lead author of *Systematic Approaches to a Successful Literature Review* by Sage Publications (Booth, Papaioannou & Sutton, 2011), and since 2004 has contributed to 18 methodological articles, two Cochrane methodological guidance documents and 16 articles applying methodology to various aspects of evidence synthesis (See **Supporting Publications** and separate *Curriculum Vitae*). From this body of work he has selected five peer-reviewed journal articles (**Papers M1-M5**), supported by five **Case Studies**, demonstrating a coherent and unique contribution to qualitative evidence synthesis:

Methodological papers

Paper M1. Booth A. "Brimful of STARLITE": toward standards for reporting literature searches. *J Med Libr Assoc.* 2006 Oct;94(4):421-9 (1st of 1 author) [28 Web of Science Citations (24/12/2012); 49 Google Scholar Citations (24/12/2012)]

Contribution: The Candidate was solely responsible for all aspects of this study including conception, design, analysis and writing up.

Paper M2. Dixon-Woods, M, Booth, A & Sutton AJ. Synthesizing qualitative research: a review of published reports *Qualitative Research* 2007 Aug 7: 375-422 (2nd of 3 authors) [28 Web of Science Citations (24/12/2012); 70 Google Scholar Citations (24/12/2012)]

Contribution: The Candidate conceived the idea of reviewing published reports in connection with **Paper M1** ("Brimful of STARLITE"). He identified all the published reports and compiled the initial bibliographic dataset. The lead author, Dr Dixon-Woods, then identified an opportunity to extend the previous analysis beyond the literature search phase to include quality assessment and synthesis and she tabulated the data from the dataset. The Candidate was involved, together with the third author Dr Sutton, in validating the extracted data. Dr Dixon-Woods then produced a first draft, to which the Candidate primarily contributed revisions, with additional comments from Dr Sutton. All three authors approved the final draft. *[Description of Contribution Confirmed by Professor Dixon-Woods]*

Paper M3. Carroll C, Booth A, Lloyd-Jones M. Should we Exclude Inadequately-reported Studies from Qualitative Systematic Reviews? An Evaluation of Sensitivity Analyses in Two Case Study Reviews, *Qualitative Health Research* 2012; 22 (10): 1425-1434. (2nd of 3 authors) [0 Web of Science Citations (24/12/2012); 2 Google Scholar Citations (24/12/2012)]

Contribution: The Candidate and Dr Carroll identified quality assessment as a key issue for qualitative syntheses. Dr Carroll conceived the idea for 'qualitative sensitivity analyses' to explore implications from the two case study reviews. The Candidate was Principal Investigator on the E-Learning case study and contributed to data analysis and interpretation. Dr Carroll performed the initial sensitivity analyses and wrote the first draft. The Candidate contributed substantively to manuscript revision, with additional comments from Dr Lloyd Jones. All three authors approved the final draft. *[Description of Contribution Confirmed by Dr Carroll]*

A 'Dual Heritage' for QES

Paper M4. Carroll C, Booth A, Cooper K. A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Medical Research Methodology*. 2011 Mar 16;11:29. (2nd of 3 authors). [10 Web of Science Citations (24/12/2012); 18 Google Scholar Citations (24/12/2012)]

Contribution: The Candidate and Dr Carroll jointly selected framework synthesis as the most appropriate methodology for the synthesis. The Candidate identified the initial framework from literature and Internet searches. Dr Carroll performed the initial data extraction with support from Dr Cooper. The Candidate validated the data extraction and refined the framework by suggesting the addition of the Transtheoretical Model. Dr Carroll and the Candidate conceived the idea for the paper and Dr Carroll produced the first draft. The candidate suggested substantive revisions with additional suggestions from Dr Cooper. All three authors approved the final draft. *[Description of Contribution Confirmed by Dr Carroll]*.

Paper M5. Booth A, Carroll C, Ilott I, Low LL & Cooper K. Desperately Seeking Dissonance: Identifying the "Disconfirming Case" in Qualitative Evidence Synthesis. *Qualitative Health Research* 2013; 23 (1): 126-141 (1st of 5 authors) [0 Web of Science Citations (24/12/2012); 0 Google Scholar Citations (24/12/2012)]

Contribution: The Candidate identified the niche for this particular study and led on production of the first draft and all subsequent revisions. The list of strategies for maximising identification of the disconfirming case was uniquely identified by the Candidate, drawing upon case studies from three component reviews. Collaborators from the three case studies verified methodological details from their respective studies and contributed additional insights to the drafts.

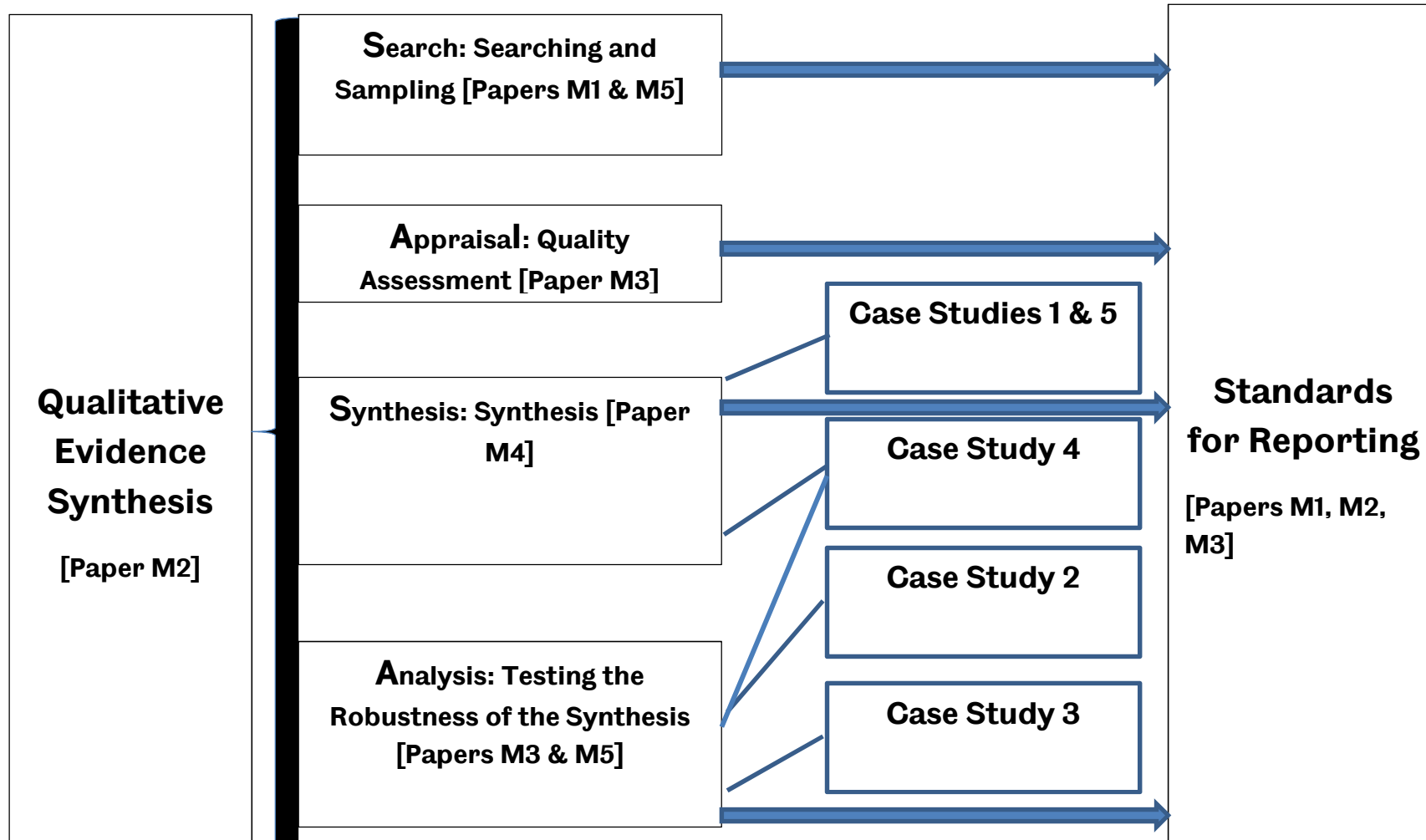
I confirm that the accompanying integrative commentary is the sole work of the candidate.

A handwritten signature in black ink, appearing to read 'A. Booth', written over a horizontal line.

27/12/2012

A 'Dual Heritage' for QES

Figure 1 - Relationship between Included Papers



A 'Dual Heritage' for QES

Case studies as methodology lead

Case Study 1 – Chemoprevention of colorectal cancer (National Institute for Health Research Health Technology Assessment Programme; 2007-2009)

Contribution: The Candidate was lead methodologist for the qualitative component of this three part technology assessment report. The Candidate's discrete contribution is identifiable in *Chapter 4 - Review of qualitative data on views and attitudes to the taking of agents that may be used for chemoprevention*. The Candidate contributed, with Dr Carroll and Dr Cooper, to data extraction, synthesis, interpretation and to all of the drafts which were managed by Dr Carroll. This Case Study contributed to **Paper M4**.

Published Outputs: Carroll et al, 2010a, 2010b; Cooper et al, 2010; Papaioannou et al, 2011; Squires et al, 2011.

Case Study 2 –Student experience of workplace-based e-learning (Higher Education Academy; 2007)

Contribution: The Candidate was Principal Investigator on this Higher Education Academy funded project. He led on the methodology of the review and contributed to data extraction and analysis. This Case Study contributed to **Papers M3 and M5**.

Published Outputs: Booth et al, 2009; Carroll et al, 2009a, 2009b, 2011; Booth 2010; Papaioannou et al, 2010; Booth 2011d.

Case Study 3 – The nurse, midwife and health visitor contribution to protocol based care (National Institute for Health Research Service Delivery and Organisation Programme; 2006-2008)

Contribution: The Candidate was the key methodologist on the qualitative synthesis component of this project. He was jointly involved with Dr Ilott in data extraction and quality assessment and produced one of the two syntheses. He contributed significantly to revision of the draft manuscripts. This Case Study contributed to **Paper M5**.

Published Outputs: Patterson et al, 2010; Ilott et al, 2006, 2010.

Case Study 4 – Group therapies for postnatal depression (National Institute for Health Research Health Technology Assessment Programme; 2007-2009)

Contribution: The Candidate was the key methodologist on the qualitative synthesis component of this project. He was jointly involved with Dr Scope in data extraction and contributed instrumentally and conceptually to synthesis, analysis and interpretation. He also contributed significantly to revision of the draft manuscripts. This Case Study informed **Papers M4 and M5**.

Published Outputs: Stevenson et al, 2010; Scope et al, 2012

Case Study 5 – The evidence base for emergency planning (National Institute for Health Research Service Delivery and Organisation Programme; 2010-2012)

Contribution: The Candidate was the key methodologist on the scoping review of published journal literature. He conceived the framework analysis approach using the FEMA framework. He was jointly involved with Drs Challen, Lee and Mr Gardois in data extraction and contributed significantly to writing and revision of the draft manuscripts. This Case Study utilised learning from **Paper M4**.

Published Outputs: Challen et al, 2012; Lee et al, 2012.

A 'Dual Heritage' for QES

Coherence of Work

Papers M1 and M2 survey existing methods of searching and qualitative evidence synthesis. They provide a platform for empirical investigations (**Papers M3-M5**) of specific aspects of quality assessment, synthesis and analysis. These investigations were informed by five exemplar case studies (**Case Studies 1-5**).

1. Background

The systematic review of effectiveness is properly regarded as a building block for evidence based healthcare and evidence based policy (Tranfield et al, 2003). While the methodological origins for the systematic review may be traced over several centuries (Petticrew, 2001) the momentum behind progress in the United Kingdom can be pinpointed to the mid-1990s (Chalmers et al, 2002). In 1994 the UK Cochrane Centre (Lefebvre, 1994) and the NHS Centre for Reviews and Dissemination (Glanville, 1994) were funded within the NHS Information Systems Strategy. In the following year the first text on systematic reviews was published by the BMJ Publishing Group (Chalmers & Altman, 1995). More significantly for the methodology community the first guidance on how to conduct systematic reviews was issued by the NHS Centre for Reviews and Dissemination one year later (Deeks et al, 1996).

Early guidance focused on effectiveness and had nothing to say about alternative models of review, addressing different study questions, diverse types of evidence or variants in study design. Methods for the synthesis of qualitative research did already exist (meta-ethnography, for example, dates from the late 1980s when two educational researchers, Noblit & Hare (1988), used this technique to synthesise school inspection reports). In a landmark study, *Adapting to and Managing Diabetes* (Paterson et al, 1998), published a decade after Noblit & Hare's book, nursing researchers applied meta-ethnography techniques to forty-three qualitative interpretive research reports. Paterson and colleagues (1998) identified balance as "the determinant metaphor of the experience of diabetes". In contrast to the "closed set" of documents used by Noblit & Hare, Paterson and colleagues harnessed meta-ethnography to systematic search procedures. They identified study reports from "six computerized data bases 1980-1996" covering nursing, social sciences, and allied health. They acknowledged the systematic review method implicitly, describing their technique as "ethnographic meta-analysis", but traced their heritage to Zhao (1991), predating systematic review procedures.

In the late 1990s the prevailing view towards incorporation of qualitative research within systematic reviews appeared to be to adopt the systematic review method and adapt it, albeit minimally, to a different 'type' of research. Teaching of systematic

A 'Dual Heritage' for QES

review methods, if acknowledging the utility of qualitative research at all, typically chose to diverge at the stage of synthesis. The implication was that all other stages of the process were common to both paradigms. In short, a systematic review was conventionally understood as characterised in **Box 1** (Cook et al, 1998).

Box 1 Characteristics of a Systematic Review (Cook et al, 1998)

The characteristics of a systematic review include:

- Often a focused clinical question
- Comprehensive sources and explicit search strategy
- Criterion-focused selection, uniformly applied
- Rigorous critical appraisal
- Quantitative summary

This Thesis challenges assumptions underlying all these characteristics. It is shaped by involvement in qualitative evidence synthesis (QES) *and* in the development of alternative review products with differing degrees of systematicity (Carroll, Cooke,...Booth et al, 2006; Wilkinson...Booth et al, 2009). Consequently, the included methodology papers are included within the pragmatic context of 'decision support' (i.e. health technology assessments, rapid evidence assessments for government departments and reviews commissioned by the National Institute for Health Research) rather than 'knowledge support' (Mays et al, 2005). Such pragmatism positions **Papers M3 to M5** as practical, opportunistic responses to issues encountered when conducting 'real world reviews' (Bambra, 2011) not as a prespecified programme of methodologically coherent sub-projects (cp. **Papers M1 & M2**).

Furthermore, largely as a consequence of the above, this Thesis favours use of the phrase 'systematic approaches to the literature' (Booth, Papaioannou & Sutton, 2011) as offering greater flexibility than the idea of a 'systematic review'. 'Systematic approaches' share the key stages of **Searching**, **Appraisal**, **Synthesis** and **Analysis**, which form the acronym and mnemonic, SALSA (Grant & Booth, 2009; Booth, Papaioannou & Sutton, 2011), and yet differ in rigour and the resource required for each stage. The fourth element, Analysis, acknowledges that bringing together and

A 'Dual Heritage' for QES

interpreting studies is often not an endpoint, but a starting point for inquiry. Methodological developments, such as the scoping review, the mapping review, and the rapid evidence assessment, (Grant & Booth, 2009) demonstrate that the SALSA elements (**Box 2**) contribute variability to all systematic approaches. In contrast the first two elements of a conventional systematic review (**Box 1**), namely existence of a study protocol and a highly focused question are not prerequisite to a 'systematic approach'. Existence of a study protocol may inhibit iterative methods of searching and synthesis. Similarly prespecification of a highly-focused question negates progressive identification and refinement of a question, characterised by grounded theory approaches from primary research (Finfgeld-Connett & Johnson, 2013).

Box 2 Characteristics of Systematic Approaches to the Literature (Booth, Papaioannou & Sutton, 2011)

Systematic approaches to the literature are characterised by SALSA:

1. Explicit methods for **S**earching
2. Explicit methods for **A**ppraisal
3. Explicit methods for **S**ynthesis of studies
4. Explicit procedures for **A**nalysis of findings

A review protocol and a review question may develop iteratively

1.1 Qualitative Systematic Reviews: same family, different heritage?

In 2001, the Author presented a paper at a Qualitative Evidence Based Practice conference, *Cochrane or Cockeyed?: How should we conduct systematic reviews of qualitative research?* (Booth, 2001). This paper provocatively challenged the default position that methods for systematic review, championed so powerfully by the Cochrane Collaboration, could be applied wholesale and uncritically to newly emerging systematic reviews of qualitative research. Opportunities to develop qualitative evidence synthesis were apparently constrained by:

the implicit assumption that systematic review methods that are conceived and developed in the context of quantitative reviews should be translated uncritically for use in qualitative reviews (Booth, 2001).

A 'Dual Heritage' for QES

In challenging this assumption Booth (2001) posited that “systematic reviews of qualitative research might acquire a methodology...more sympathetic to the paradigm within which they are conducted”. He argued that established techniques from primary qualitative research, for example theoretical saturation, may contribute to a more appropriate methodology toolkit. This stance represented an attempt to resist a tendency subsequently identified as “mission drift”:

A mistake is often made, however, in transposing methods best suited to systematic review of quantitative studies into qualitative ones...producing a kind of 'mission drift' in many qualitative 'systematic' research reviews. (Jones, 2004)

Among exemplars of such mission drift Jones singled out:

Check-lists, 'standards', matrices, 'hierarchies of evidence' and other terminology borrowed from the arsenal of the quantitative camp pepper qualitative ground like so many cluster bombs... (Jones, 2004).

1.2 Towards a 'Dual Heritage' model of Qualitative Evidence Synthesis

Tensions between the respective heritages of qualitative research and systematic reviews had surfaced in a methodological review of *Qualitative research methods in health technology assessment* (Murphy et al, 1998). The review team expressed objections to the systematic review method, arguing that such a positivistic, hypothetico-deductive approach was anathema to the qualitative research paradigm. While the team's critique did not argue against the usefulness of systematic review methods *per se*, it did attempt to specify prerequisites to successful use:

The topic being studied must be in a state of... 'normal science' where there is a high degree of consensus on the definition of problems and methods, where there are accepted means of defining these operationally which lead to a standard use of keywords and where the results come in forms that can be treated as equivalent or converted into a common currency (Murphy et al, 1998).

Occasionally this caricature of systematic review methods betrays preconceptions that this methodology is essentially reductionist:

...all professional judgements are eliminated by objective scoring systems that allow all results to be fed into a single matrix, which can then be analysed by impersonal means. This approach works well under certain limiting conditions. (Murphy et al, 1998)

A 'Dual Heritage' for QES

In contrast to such a dichotomy, this Thesis relates to a 'dual heritage' for QES methodology. 'Dual heritage' literally refers to having parents from different cultural (and/or ethnic) backgrounds. By extension it is used metaphorically to indicate the rich diversity accessed by QES in drawing upon the cultures, or research traditions, of *both* qualitative research and systematic review. We must firstly acknowledge that qualitative research is, itself, constituted of multiple cultures and traditions, some almost as distant from each other as quantitative research is from qualitative research. However a pragmatic focus permits dipping judiciously into this methodological 'gene pool', the entire qualitative 'genome', so to speak. While metaphorical usage of 'dual heritage' is uncommon it is not without precedent (e.g. Kvan, 2004). This Thesis treads a conciliatory path by making explicit that some appropriate QES techniques derive from the conventional model of the systematic review process while others originate from the methods of primary qualitative research. Thomas & Harden (2008) acknowledge this dual heritage:

When we started...reviews which included qualitative research in 1999, there was very little published material that described methods for synthesising this type of research. We therefore experimented with a variety of techniques borrowed from standard systematic review methods and methods for analysing primary qualitative research.

However the concept of 'dual heritage', used in this Thesis, legitimises these respective legacies beyond the idea of simply 'borrowing'. Indeed this dual heritage had earlier been articulated by the same team:

We combined conventional systematic review principles and methods with more novel ones developed in the course of the review series. Using the techniques and terminology usually associated with qualitative analysis of primary research data helped us to be systematic and explicit (Harden et al, 2004).

Instead of contrasting two antagonistic heritages, this observation suggests that primary research methods can synergistically contribute to the systematic review requirement to be systematic and explicit. By way of contrast, Major & Savin-Baden (2010), from a social science research tradition, underplay the systematic review heritage, stating that "a qualitative research synthesis, then, uses qualitative methods to synthesize existing qualitative studies to construct greater meaning through an interpretive process". Perversely Major & Savin-Baden (2010) further specify a need

A 'Dual Heritage' for QES

to focus “on a narrow research question”; a legacy of the systematic review heritage. Problematically, the notion of a ‘narrow’ review question fails to recognise that definition of a research question, in the context of QES, may be an early review *outcome*, not necessarily a starting point.

While the richness of a dual heritage constitutes a considerable asset, it has understandably contributed to confusion regarding the ‘identity’ of the synthesis product. Some methods of synthesis naturally gravitate towards one branch of the heritage (e.g. meta-aggregation as practised by the Joanna Briggs Institute engages overtly with the systematic review tradition). Perversely meta-ethnography, notwithstanding its proximity to grounded theory-based approaches, demonstrates a strong kinship with conventional systematic review procedures, although never conceived as such.

Counterposing the two origins of the heritage antagonistically can result in the misplaced assumption that either variant is sub-standard, or simply “wrong”. For example Murphy and colleagues (1998) range their “Nottingham model” *against* the “York CRD model”.

our approach is essentially a qualitative one which exhibits the strategies of the kind of qualitative research that we ourselves do. The ‘Nottingham Model’, if we may call it that, uses a process of analytic induction.

Murphy and colleagues (1998) cite the usefulness of candidate procedures from the qualitative heritage such as constant comparison to compare “different authors’ approaches to the same issues” and deviant case analysis in looking for “authors who offer different perspectives”. Furthermore they describe how reviewers are able “to establish when there is no more to say on an issue (theoretical saturation)” (Murphy et al, 1998). Their methodology review, with qualitative research as its subject matter, was the first to employ the terminology of primary qualitative research to describe the rationale and procedures of systematic review.

In situations where both heritages offer viable, yet contrasting, methods, a reviewer may be confused at seemingly contradictory advice. Ideally the stimulus afforded by a

A 'Dual Heritage' for QES

'dual heritage' model should lead a reviewer towards methodological exploration and innovation, yielding solutions that satisfy the rigour required by review methods, coupled with sensitivity to the qualitative paradigm. **Table 1** attempts to document the substantive contribution of each heritage to a QES toolkit, although simplifying the continuum by presenting a dialectic. This Thesis builds upon tensions and creative energies experienced from working within this 'dual heritage', as evidenced in **Papers M1-M5**, to identify its ongoing impact upon the key review stages of study identification (searching), quality assessment (critical appraisal) and synthesis.

A ‘Dual Heritage’ for QES

Table 1 - Potential Contributions to Qualitative Evidence Synthesis of the respective heritages

Review Phase	Systematic Review Heritage	Primary Qualitative Research Heritage
Context for question (Oliver et al, 2005)	Strips away context, subsequently revisited as generalisability	Explores context, studies are situated
Review question (Dixon-Woods, Bonas...Booth et al, 2006)	Starts from fixed, predetermined question, using PICO format	Treats question as negotiable, emerging, “Compass rather than Anchor” (Eakin & Mykhalovskiy, 2003)
Sampling (Suri, 2011)	Employs comprehensive sampling	Draws upon purposive, theoretical or maximum variability sampling (Suri, 2011)
Search strategy (Noyes, Popay...Booth et al, 2008)	Endeavours to be exhaustive	Continues until theoretical saturation is reached (Brunton et al, 2012)
Search process (Pearson et al, 2011)	Is prescribed by a protocol	Is viewed as iterative (Brunton et al, 2012)
Quality assessment (Manning, 2011)	Involves application of uniform criteria	Treats quality as contested, both as a whole and in terms of appropriateness for particular types of qualitative research
Assessment process (Hannes, 2011)	Used to include/exclude	Used to moderate interpretations
Synthesis approach (Dixon-Woods, Bonas...Booth et al, 2006)	May be characterised as aggregative	May be perceived as interpretative/configurative (Gough et al, 2012a; 2012b)
Synthesis methods (Noyes & Lewin 2011b)	Employs narrative synthesis (“epidemiology” of studies)	Uses framework analysis, thematic analysis (Barnett-Page & Thomas, 2009)
Analysis (Gough et al, 2012b)	Maps study elements	Explains or applies existing (or even creates new) constructs
Sensitivity analysis (Harden, 2008)	Explores differences in Population, Intervention, methods of outcome measurement and study quality	Explores differences in context, thickness of detail, conceptual richness
Approach to heterogeneity (Candy et al, 2011)	Seeks to establish commonality, “averaging effect”	Explores context as an explanation for difference (Hannes & Harden, 2012)
Documentation (Booth, Papaioannou & Sutton, 2011)	Utilises PRISMA structure and flow diagram	Utilises diagrams, schema, conceptual models etc

2. Context for the Featured Papers

In 2005 the Economic and Social Research Council (ESRC)-funded a multidisciplinary project team to examine “How can systematic reviews incorporate qualitative research?” This project articulated a coherent thread that runs through all the featured papers, namely: “a need to explore whether the template offered by conventional systematic review methodology can comfortably accommodate qualitative research” (Booth, 2001). In recognition of this concern, our ESRC team chose to problematize the apparently irreconcilable tensions from bringing together the two heritages:

Throughout our project we experienced difficulty with matching the tasks and epistemological assumptions associated with qualitative research with the template offered by conventional systematic review methodology (Dixon-Woods, Bonas...Booth et al, 2006)

We further reported that we had found it:

inappropriate or impossible to specify a clearly focused review question; to use completely reproducible and transparent search and selection strategies; or to construct an inherently reproducible synthesis (Dixon-Woods, Bonas...Booth et al, 2006)

The main output from the ESRC project, a nine-author methodological paper in *Qualitative Research* (Dixon-Woods, Bonas...Booth et al, 2006), has subsequently been cited 69 times on Web of Science and 150 times on Google Scholar (24/12/2012) and serves as a backdrop to subsequent work in this Thesis. Involvement in the ESRC Project yielded opportunities to explore all stages of the review process, including quality assessment and synthesis. To orientate the ESRC Project, Booth created a dataset of existing published qualitative systematic reviews to support two analyses of methodological characteristics (**Papers M1 & M2**).

Paper M1. "Brimful of STARLITE": toward standards for reporting literature searches.

Paper M2. Synthesizing qualitative research: a review of published reports.

Participation in a Cochrane Collaboration-funded Methods Summit [Adelaide, Australia, 2008] further helped to shape QES methodology, informing production of supplemental methodology guidance (Harris 2011; Booth, 2011a; Hannes, 2011; Noyes & Lewin, 2011a, 2011b) and two subsequent books (Booth, Papaioannou & Sutton,

A 'Dual Heritage' for QES

2011; Hannes & Lockwood, 2012). Methodological issues identified from the Summit by the Cochrane Qualitative Methods Group Co-convenors, subsequently informed three sub-projects (**Papers M3-M5**):

Paper M3. Should we exclude Inadequately Reported Studies From Qualitative Systematic Reviews? An Evaluation of Sensitivity Analyses in Two Case Study Reviews.

Paper M4. A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents.

Paper M5. Desperately Seeking Dissonance: Identifying the "Disconfirming Case" in Qualitative Evidence Synthesis.

All the above methodology papers, published in highly regarded journals, utilise empirical findings from a test bed of funded reviews (**Case Studies 1-5**) (**Figure 1**):

Case Study 1 – Chemoprevention of colorectal cancer (National Institute for Health Research Health Technology Assessment Programme; 2007-2009)

Case Study 2 – Student experience of workplace-based e-learning (Higher Education Academy; 2007)

Case Study 3 – The nurse, midwife and health visitor contribution to protocol based care (National Institute for Health Research Service Delivery and Organisation Programme; 2006-2008)

Case Study 4 – Group therapies for postnatal depression (National Institute for Health Research Health Technology Assessment Programme; 2007-2009)

Case Study 5 – The evidence base for emergency planning (National Institute for Health Research Service Delivery and Organisation Programme; 2010-2012)

Research methods for the chosen outputs (**Papers M1-M5**) utilise three approaches previously showcased in the *Cochrane or Cockeyed* paper (Booth, 2001):

1. systematic review of the methodology literature,
2. analysis of published case reports of literature review projects and
3. empirical exploration of methodological issues within ongoing systematic review case studies.

3. On Searching and Sampling

M1. "Brimful of STARLITE": toward standards for reporting literature searches.

M2. Synthesizing qualitative research: a review of published reports.

M5. Desperately Seeking Dissonance: Identifying the "Disconfirming Case" in Qualitative Evidence Synthesis.

Selection of an appropriate sample of *participants* is fundamental to either quantitative or qualitative primary research (Onwuegbuzie & Collins, 2007). Similarly selection of an appropriate sample of *papers* is essential for successful evidence synthesis. Quantitative primary research (Cooper, 2007), and indeed quantitative synthesis, seeks a representative sample, to minimise bias and to permit subsequent generalisation. Unfortunately, within the context of systematic reviews, this objective is often interpreted simplistically as requiring a comprehensive and exhaustive sample. As a consequence, thinking around systematic identification of an *appropriate* sample of papers is comparatively underspecified. It is only recently that commentators have acknowledged that the key consideration is the appropriateness of a sample, not necessarily its exhaustivity. For example Brunton and colleagues (2012) observe that:

Exhaustive searching is improbable, because the total universe of potentially relevant literature is unknown....the obligation on reviewers is to plan a thoughtful and clearly described plan [*sic*] to locate the sample of studies most likely to answer their research question[s] reliably.

The persistence of the "myth" of exhaustivity (Brunton et al, 2012) has held undesirable ramifications for the emergence of QES. The benchmark of a comprehensive sampling frame has persisted despite the obvious utility of purposive and theoretical sampling approaches, as practised by primary qualitative research. Indeed, the idea of adding additional electronic databases to a search protocol, (i.e. to search for more of the same) (Subirana et al, 2005), runs counter to the strategy of seeking to diversify a sample (i.e. purposively to move on to different, more productive lines of inquiry). The reviewer is, therefore, "not concerned with piling up examples of the same finding, but in identifying studies that contain new conceptualisations of the phenomena of interest" (Brunton et al, 2012).

A 'Dual Heritage' for QES

The Author was one of the earliest to challenge the implicit synonymy between the requirements that reviews be systematic and that they be comprehensive:

[Systematic] reviews....make much of the virtue of the comprehensiveness of the data sources that have been searched..... Is it necessary for a qualitative review to be equally comprehensive...? (Booth, 2001)

There are good methodological reasons for attempting to search comprehensively in quantitative systematic reviews, not least as a protection against systematic errors such as publication bias (Lefebvre et al, 2011; Brunton et al, 2012). The analogy here is with the need for systematic representative sampling in quantitative population surveys. However qualitative data collection is characterised, not by “statistical representativeness” but by “systematic non-probabilistic sampling” (Mays & Pope, 1996).

Papers M1 and M2 represent systematic analyses of published examples of qualitative systematic reviews. **Paper M1** focused on the conduct and reporting of search strategies and concluded with a combined acronym/mnemonic, STARLITE, designed to identify the essential elements to be reported in future examples of QES. **Paper M2** broadened analysis beyond the search process to examine all stages of the review process including quality assessment and synthesis. When it was written (early 2005), commentaries on searching for qualitative research studies within the systematic review heritage typically started from an assumption that comprehensiveness is similarly prerequisite for QES. However commentators on QES methodologies in general, particularly those with an interpretive intent, were starting to question this apparent necessity (Weed, 2005; Pawson et al, 2005). For example guidance from the Centre for Reviews and Dissemination (CRD), acknowledges:

no consensus as to whether the searches undertaken to identify qualitative studies need to be as comprehensive...as those undertaken to identify quantitative studies, although they should be as systematic, explicit and reproducible as possible (CRD, 2008).

Paper M1 sought, within its overall objectives, to establish variation in the methods of sampling used by QES published between 1988 and December 2004. The study found that forty-four of the 65 included reviews (68%) reported sufficient details of their

A 'Dual Heritage' for QES

search methods to permit identification of their sampling strategy. Thirty seven reviews employed the comprehensive sampling strategies that characterise the systematic review heritage. In contrast six used purposive sampling and one used opportunistic sampling, both characteristic of the primary qualitative heritage. Although the sizeable majority favoured comprehensive search strategies this significant methodological confusion justified investigation of appropriate sampling search strategies (**Paper M1**). Booth (2010) subsequently reviewed alternatives for determining adequacy of search strategies asking "How much searching is enough?".

The remaining twenty-one studies did not report any elements of their search strategies. Based on these findings, **Paper M1** advocated a structure for reporting literature searches under the acronym/mnemonic STARLITE. The S of STARLITE highlights **S**ampling strategies as a possible source of variation. Henceforth one need no longer assume that all QES have employed a comprehensive sampling strategy. STARLITE is recognized by the EQUATOR Network (Simera & Altman, 2009) as "the first attempts to harmonise reporting of a particular aspect of research, such as the search strategy for systematic reviews" (Simera et al, 2008). Sampson et al, (2008) conducted a systematic review that identified eighteen instruments for reporting search strategies. Of these 18 instruments, STARLITE was the only guideline to suggest inclusion of sampling strategy as a significant item. While this observation can be explained, in part, by the focus of other instruments on quantitative systematic reviews it is significant that STARLITE anticipated requirements for an increasingly more versatile range of review products, such as mapping and scoping reviews, meta-narratives, critical interpretive syntheses and realist syntheses (Urquhart, 2010).

Paper M2 augmented the previous analysis (**Paper M1**) to not only include the search phase but also quality assessment and synthesis. It indicated that difficulties in identifying primary qualitative research papers were shared when identifying QES (Shaw, Booth et al., 2004). The review team encountered:

'false negatives' (publications that fail to be identified as reviews of qualitative research) as well as 'false positives' (some reviews appear to be 'qualitative systematic reviews' but are simply using the term 'qualitative' to indicate that

A 'Dual Heritage' for QES

they have not used a formal method of quantitative synthesis such as meta-analysis) [**Paper M2**]

Practical difficulties in identifying QES continue to inform search strategies by which the Author populates the study register of the Cochrane Collaboration Qualitative Methods Group with methodological articles and published examples.

By confirming limitations of protocol-driven approaches to searching (Greenhalgh and Peacock, 2005; Pearson et al, 2011) and limitations of aiming for 'comprehensiveness' (Dixon-Woods, Cavers et al., 2006), **Papers M1 and M2** stimulated an investigation of the value of supplementary search techniques (Papaioannou,...Booth et al, 2010). **Paper M2** also argued that methods for searching must be systematic to forestall "criticisms that choice of studies was idiosyncratic and capricious" and identified instances of suboptimal search strategies. For example, "many keyword search strategies revealed...a lack of clarity regarding use of free text terms or approved subject headings". Many reviewers failed "to use methodological terms in addition to subject-specific terms (e.g. the term "qualitative") of the type that would increase specificity in retrieving qualitative studies". The prevalence of suboptimal searching reinforced "the benefits of including...an information specialist skilled in searching and retrieval techniques", as advanced in previous papers (Beverley, Booth & Bath, 2003; Wilkinson,...Booth et al, 2009).

Paper M2 has been cited to justify specific literature search methodologies using extensive lists of keywords and subject terms, and judicious purposive sampling (Sampson et al, 2011). This first-ever survey of published QES has been replicated and updated for the period 2005 to 2008 (Hannes & Macaitis, 2012). Hannes & Macaitis used search methods described in **Paper M1** to replicate **Paper M2**. They document a move away from protocol based bibliographic database search methods towards increased use of "supplementary search strategies". They also acknowledge debate concerning the desirability of comprehensive and exhaustive search methods, by agreeing that search strategies aiming to identify qualitative research need to be "systematic and explicit". At the same time they acknowledge that "the need for

A 'Dual Heritage' for QES

comprehensive, exhaustive searches in qualitative research is questioned". They observe that

Theoretical and purposive sampling might be a feasible alternative as long as the 'picture' from the studies that have been retrieved incorporate all likely insights (Hannes & Macaitis, 2012).

Hannes & Macaitis (2012) conclude their update of **Paper M2** by citing Cochrane guidance (Noyes,...Booth et al., 2008) supporting the need to determine "when and how these contrasting sampling philosophies are to be used appropriately" (Hannes & Macaitis, 2012)

The need to match a search strategy to the most appropriate sampling approach (**Table 2**) is preferable to any blanket approach advocating comprehensive sampling (**Papers M1 & M2**). Recent years mark recognition that the *appropriateness* of sampling, not comprehensiveness, provides the quality marker for a well-conducted QES (Petticrew & Roberts, 2006).

A 'Dual Heritage' for QES

Table 2 - Synthesis Methods with Appropriate Sampling Methods

Synthesis Method	Sampling Method
Critical Interpretive Synthesis	Purposive Sampling (Dixon-Woods et al, 2006)
Meta-Ethnography	Purposive Sampling (Doyle, 2003)
Meta-Interpretation	Maximal Divergent Sampling (Corbin-Staton, 2009)
Meta-Narrative Synthesis	Purposive Sampling of key papers within different research 'traditions' (Barnett-Page & Thomas (2009)
Qualitative meta-synthesis	Comprehensive (representative) Sampling (Paterson et al, 2001)
Realist Synthesis	Comprehensive Sampling (Brunton et al, 2010); Purposive Sampling (Pawson, 2006c); Snowball Sampling (Pawson et al, 2004)
Scoping Review	Random Sampling (Brunton et al, 2010)

Paper M5 further challenges the appropriateness of a “gold standard” comprehensive search strategy, this time within a broader and more sophisticated awareness of the circumstances (and review types) under which particular sampling methods might be appropriate:

there is increasing recognition that, for a qualitative evidence synthesis, it is more critical that a search strategy is selected to match the intended purpose of the review. **[Paper M5]**

Although specific methods, such as critical interpretive synthesis and realist synthesis, may particularly benefit from purposive sampling methods advantages may accrue more widely:

Other methods too might derive value from the diversity, rather than the comprehensiveness, of their sampling technique. Purposive or theoretical sampling, in a quest to achieve theoretical saturation, might increase the likelihood that reviewers retrieve disconfirming cases. **[Paper M5]**

A 'Dual Heritage' for QES

For example:

A review team might use literature searches to operationalize maximum variation sampling by accessing disciplines or schools of thought that emphasize diversity and dissonance [**Paper M5**]

This latter observation implicitly acknowledges a dual heritage with primary research methods. It confirms an earlier comment that:

other principles from primary qualitative research methods may also be 'borrowed' such as deliberately seeking studies which might act as negative cases, aiming for maximum variability and, in essence, designing the resulting set of studies to be heterogeneous, in some ways, instead of the homogeneity that is often the aim in statistical meta-analyses (Thomas & Harden, 2008),

Major & Savin-Baden (2011) explicitly acknowledge this useful distinction between the purpose of the review and its subsequent sampling strategy (**Paper M1**). They advance understanding of those circumstances under which different types of sampling are appropriate. For example, comprehensive sampling is most appropriate in breaking larger units down into their component parts or variables whereas interpretation of meanings across primary studies requires purposive sampling. Finally constructing new meaning from existing evidence may well require purposeful sampling until theoretical saturation is reached.

This widespread recognition of the need to match search techniques to sampling requirements represents a valuable way for progressing the methodologies of study identification. Future research directions are suggested by a paper that specifically examines sampling and study selection as an issue for QES (Suri, 2011). Booth, Harris & Hannes (2011) have subsequently explored the implications of sampling methods for a selection of search techniques at a workshop at the Madrid Cochrane Colloquium.

While the imperative to devise alternative methods of sampling is frequently driven by pragmatic considerations, such as "where the number of available studies may be simply too large to work through" (Centre for Reviews and Dissemination, 2008), methodologically it presents an opportunity to develop methods congruent with the primary qualitative research heritage. Purposive or theoretical sampling may allow

A 'Dual Heritage' for QES

the reviewer to select papers for “inclusion on the basis of particular criteria such as rich description or conceptual clarity”. Already purposive sampling can be detected in reviews of caring (Finfgeld, 2008) and access to health care (Dixon Woods, Cavers et al, 2006). Doyle (2003) advocates the use of purposive sampling in meta-ethnography:

like meta-analysis, meta-ethnography utilizes multiple empirical studies but, unlike meta-analysis, the sample is purposive rather than exhaustive because the purpose is interpretive explanation and not prediction.

Such observations lead Thomas & Harden (2008) to suggest that:

it may not be necessary to locate every available study because, for example, the results of a conceptual synthesis will not change if ten rather than five studies contain the same concept, but will depend on the range of concepts found in the studies, their context, and whether they are in agreement or not.

They echo Booth (2001) in suggesting that:

principles such as... ‘conceptual saturation’ might be more appropriate when planning a search strategy for qualitative research, although it is not yet clear how these principles can be applied in practice.

Random sampling is more controversial and probably faces a more difficult ride, given that its methods satisfy the principal objections of neither the quantitative nor qualitative review community. Nevertheless, random sampling may have a place in scoping or mapping reviews within the wider family of systematic approaches to reviewing the literature (Booth, Papaioannou & Sutton, 2011). Indeed Brunton and colleagues (2012) describe their use of random sampling in an earlier research synthesis on women’s experiences of becoming a mother (Brunton et al, 2010). They chose a random sample of 10% to match “the team’s capacity to screen in the available time” (Brunton et al, 2012).

3.1 Limitations of Papers M1 and M2

In retrospect, **Paper M1** was limited, by both time and resource constraints, from constituting a formal methodology systematic review, being described, instead, as a “survey of the literature”. As a consequence its main weakness lay in the Inclusion Criteria being implemented by a single reviewer and, therefore, open to potential subjectivity (Sandelowski, 2007). Sandelowski’s response typifies difficulties in characterising such syntheses, particularly in hybrid papers illustrating methodology

A 'Dual Heritage' for QES

through a case study of a review (Booth, 2007). Use of more explicit criteria, as implemented for **Paper M2**, would have forestalled such concerns. A secondary observation, that authors are constrained in reporting by editors and publishers, highlights that any analysis of methodology is constrained by relying on what is reported, not necessarily what has been executed (Sandelowski, 2007).

As a consequence **Paper M2** constituted a more robust investigation than **Paper M1**, being strengthened by explicit inclusion criteria and by using multiple reviewers in validating included studies. While providing a cross-sectional frame of reference for ongoing ESRC project work, and contributing to ongoing sensitization to QES methodology, the primary function of **Paper M2** was in diagnosing contemporary methodology challenges. In common with most methodological literature surveys **Paper M2** offered little in the way of methodological “solutions” although it did identify promising avenues for investigation. Indeed, at the time of publication, the ESRC project team had started to undertake empirical work in highlighted areas and had identified further research priorities in associated areas.

3.2 Summary

Papers M1 and M2 both served to map and document variation in QES practice in study identification. Furthermore both papers revealed priorities for subsequent research, research taken forward by other papers in this portfolio and by other research teams within the QES community. **Papers M1 and M2** made a significant contribution in acknowledging appropriate variations in sampling strategy; a variation that needs to keep pace with innovative methods of review and synthesis. In redefining “appropriate” methods of searching and sampling, these papers opened up the potential QES toolkit to incorporate procedures derived from the heritage of primary qualitative research.

4. On Quality Assessment

Paper M2. Synthesizing qualitative research: a review of published reports

Paper M3. Should we Exclude Inadequately Reported Studies from Qualitative Systematic Reviews? An Evaluation of Sensitivity Analyses in Two Case Study Reviews.

Of the various review phases examined by **Paper M2** perhaps quality assessment demonstrates the greatest variability and unevenness in methods and in the quality of reporting. In this context quality assessment describes “assessment of the quality of studies to be included” (Wells & Littell, 2009). It is argued, in the context of systematic reviews more generally, that “if the ‘raw material’ is flawed, then the conclusions of systematic reviews cannot be trusted” (Jüni et al., 2001). With little evidence of apparent consensus (**Paper M2**) around quality assessment (appraisal) of studies, the review team reported collective disappointment at:

the failure in most papers...to incorporate judgements of quality in any explicit way in the synthesis...the quality appraisal appeared to be simply a procedural step with little significance for the eventual synthesis or its conclusions [**Paper M2**]

Key issues relate to the relative importance of “maximizing the conceptual yield of included papers rather than determining the robustness of the study design” (**Paper M2**). Baxter et al (2010) have reaffirmed this tension within the context of reviews of public health evidence:

critical appraisals of the type used in quantitative synthesis are less appropriate for reviews of qualitative evidence where “the conceptual yield of included papers” is more important than the robustness of the study design.

The implications that a poor quality paper may make a potentially valuable conceptual contribution, embodied in the memorable expression “Digging for Nuggets: How ‘Bad’ Research Can Yield ‘Good’ Evidence” (Pawson, 2006a), has directly informed research investigating the utility of “qualitative sensitivity analyses” (**Paper M3**). This tension between “the procedural defects of individual papers”, the legacy of systematic review methods, and approaches, such as a meta-narrative review or a critical interpretive synthesis, that take “the whole of a body of literature as its object of inquiry” (Dixon-Woods, Cavers et al., 2006; Greenhalgh et al., 2005) presaged recent methodological developments.

A 'Dual Heritage' for QES

The role of quality assessment in qualitative research is frequently contested with little consensus on whether it should be done at all and, even if so, how this might best be performed (Dixon-Woods, Bonas...Booth et al 2006; Garside et al, 2008). In the specific context of QES, Dixon-Woods, Bonas...Booth et al (2006) argue that:

Whether quality appraisal of qualitative research should be undertaken for purposes of a systematic review is a matter of some debate.

Such a debate becomes more intense given the added complexity of using such assessment as a mechanism for either including or excluding a study from a synthesis, particularly given the implications that exclusion implies. In this context use of a formal published checklist to justify summative verdicts might be seen as political expediency rather than methodological necessity. The ESRC project explored a key controversy in quality assessment, namely whether checklist-based approaches, as used in conventional systematic reviews, are appropriate in QES (Chapple & Rogers, 1998; Barbour, 2001; Power, 2001; Williams, 2001; Dixon-Woods et al, 2004). Dixon-Woods, Sutton...Booth et al (2007) highlighted an important distinction between the epistemological and technical functions of quality assessment, a distinction later embodied in guidance emerging from the Cochrane Collaboration Qualitative Methods Group Methods summit in Adelaide (Hannes, 2011).

However a further associated issue, for which only minimal evidence exists, relates to the implications of exclusion of poor quality studies. The prescription, from systematic review methodology, that each paper requires a comparable expenditure of resource, in terms of retrieval, quality assessment and subsequent synthesis, irrespective of its eventual contribution to the final product, had already emerged as a concern for pragmatic reviewers employed in our institution. Could we identify papers that were unlikely to contribute to a final synthesis?

The origins for this debate date from when Thomas et al (2004) found, in their systematic review of healthy eating among children, that five of eight included qualitative studies met nine or more of 12 quality criteria. The remaining three met six or fewer criteria. In a subsequent "sensitivity analysis" they found that the "results

A 'Dual Heritage' for QES

of these three studies did not contradict those from studies of a higher quality" (Thomas et al, 2004). They concluded that the "synthesis would have come to the same conclusions with or without their inclusion". On the basis of this experience they resolved that they would, in future "exclude poorer quality studies from the synthesis". They stated that they were "conducting methodological work to assess the impact this has on the findings of the review".

Although data from this methodological work have not been formally published, findings can be reconstructed from a subsequent report (Thomas & Harden, 2007) and associated article (Thomas & Harden, 2008). The latter documents that:

since there is little empirical evidence on which to base decisions for excluding studies based on quality assessment,...we excluded only studies which had significant flaws and used 'sensitivity analyses' to assess the possible impact of study quality on the review's findings (Thomas & Harden, 2008)

Thomas and Harden (2008) reported having assessed the 'quality' of their studies "with regard to the degree to which they represented the views of their participants". This approach positions quality within a "fitness for purpose" context for the review itself rather than within the original context of the primary studies. Such a pragmatic approach captures those elements of importance to a specific review question by conflating rigour and relevance. However, given that the overarching purpose of quality assessment in their review was to moderate review findings rather than to exclude studies, this can be deemed a low risk strategy. Thomas & Harden (2008) cite as justification for including all studies, regardless of quality, that "there are no accepted – or empirically tested – methods for excluding qualitative studies from syntheses on the basis of their quality". After completing their thematic synthesis they attempted to isolate the relative contributions of the included studies to their final analysis and recommendations. They concluded that "the poorer quality studies contributed comparatively little to the synthesis and did not contain many unique themes; the better studies, on the other hand, appeared to have more developed analyses and contributed most to the synthesis".

A 'Dual Heritage' for QES

Noyes & Popay (2007) sought to relate “technical quality” and “epistemological and theoretical” richness to “thickness” (as opposed to “thinness”) (after Popay et al, 1998). They observed that studies with ‘thin’ description offer “little, if any, explanatory insights and no opportunity for generalizing”. In contrast those employing ‘thick’ description hold “greater potential for explanation and generalization to other settings and/or social groups”. As a consequence they flagged concerns regarding “focus groups and in-depth interviews, which may be more amenable to technical appraisal and meet technical quality criteria, even though they can be devoid of theory [and]...may not be of particularly good quality in terms of their ability to speak to wider issues/relevance” (Noyes & Popay, 2007). They conclude that “if the argument prevails that some quality appraisal is necessary, the problem still remains as to how this should be undertaken”.

Harden (2008) further describes experience from sensitivity analyses based on 62 primary studies from *five reviews*. *She suggests that studies judged of low quality contributed little to the overall review findings*. In commenting on these findings the CRD guidance concludes that they are “consistent with the more descriptive accounts offered about study quality and overall contribution to synthesis” (Miller et al, 2007; Noyes & Popay, 2007; Atkins et al, 2008). Better quality studies appear to make stronger contributions to the synthesis (Miller et al, 2007; Atkins et al, 2008). Conversely weaker studies appear to contribute nothing substantially different from the stronger studies (Noyes & Popay, 2007). Harden (2008) echoes Noyes and Popay (2007) in concluding that high quality studies which display both conceptual depth and rich description appear crucial. Garside (2008) extends our understanding by observing that, *for both meta-ethnography and meta-study*, “the most conceptually developed study report contributed most to the review”. This observation helps to disentangle the quality assessment and synthesis phases, suggesting that, while reviewers may consider both important, each process may operate independently of the other. However it raises further methodological challenges as to how a review team might operationalize “conceptual richness” and “thickness of description” consistently and objectively.

A 'Dual Heritage' for QES

Paper M2 has been cited to justify inclusion of all studies, irrespective of quality (Boeije et al, 2011). The paper that replicates and updates the literature survey methods of **Paper M2** for the period 2005 to 2008 demonstrates that

authors have begun to consider quality appraisal as an important procedural step that can have an impact on the final result of a synthesis (Hannes & Macaitis, 2012).

4.1 Role of the Case Studies

Paper M2 identified a research agenda, subsequently addressed by **Paper M3**, that required additional case studies. Two QES involving Carroll, with Booth (**Case Study 1**) and Lloyd-Jones (Owen et al, 2010; Carroll, Lloyd-Jones et al, 2012) as the respective co-reviewers, offered case studies with which to explore the implications of exclusion on the grounds of quality. Essentially, the team followed Harden (2008) and others (Dixon-Woods, Bonas...Booth et al, 2006; Hannes, 2011; Marshall et al, 2012) in conceiving this process as a form of "qualitative sensitivity analysis".

Paper M3 extended existing criteria used to assess the quality of reporting of studies in these two QES, excluding studies deemed to be "inadequately reported" from the subsequent analysis. The impact of exclusion was tested quantitatively and qualitatively in respect of the review findings. Exclusion of "inadequately reported studies" had no meaningful effect on the synthesis (**Paper M3**).

Although methodological findings from **Paper M3** echo those from the previously cited papers, closer examination reveals additional nuances. First, **Paper M3** identified important differences according to the different bodies of literature and, indeed, research traditions included in a review (**Paper M3**). Differential expectations in terms of the required quality and/or conceptual richness and/or "thickness of description" of reports may have differential effects on findings for different groups. For example, **Paper M3** demonstrated that publishing conventions that favour case reports in nursing journals may result in such studies performing less satisfactorily against checklists or yielding less conceptual richness than those from other disciplines (**Case Study 2**). This finding was subsequently observed in QES in protocol based care (**Case Study 3**) and postnatal depression (**Case Study**

A 'Dual Heritage' for QES

4). Exclusion of reports from the nursing literature on the grounds of technical quality may, in actuality, not impair the resultant conceptual framework too severely. However, such exclusion may limit the evidence underpinning findings that apply to this specific sub-population i.e. nurses. This situation, observed where nurses were a sub-population in the e-learning review, had even more profound implications in the protocol-based care synthesis (**Case Study 3**) where the review focused on the “nursing contribution” and in the postnatal depression synthesis (**Case Study 4**) where nurses were one of the foremost groups delivering group therapy. In practical terms a reviewer might find it helpful to reflect whether the implications of exclusion may differ according to whether the review topic is populated by a homogeneous body of literature within a single research tradition or from the perspective of a single profession, or conversely, by a heterogeneous assemblage of several bodies of literature across multiple traditions or capturing multiple perspectives. Related concerns pertain to the trade-off between historical insights from chronologically older reports and any corresponding degradation in reporting quality (Martin Hilber et al, 2012).

However a reviewer must recognise that judgements on heterogeneity are heavily dependent upon arbitrary “lumping” or “splitting” of a review question (Weir et al, 2012). To illustrate, Garside and colleagues (2008) identified congruity between a lay view of treatment of heavy menstrual bleeding (within their review) and a professional view of the same phenomena (outwith the same review). The degree of heterogeneity/homogeneity within a review relates to the phenomenon being explored and not simply how it is operationalized within a review question.

How might a review team operationalize a qualitative judgement on heterogeneity and homogeneity of the **body of literature** (in addition to conventional systematic review considerations around heterogeneity of **Population, Intervention, Comparability, Outcomes and Study Types (PICOS)**)? QES reviewers do not seem to have explored such heterogeneity using formalised procedures. An unexplored avenue for investigating heterogeneity relates to the degree of co-citation across included studies.

In writing **Paper M3** (and subsequent work on **Paper M4**) the Author strongly advocated qualitative sensitivity analysis as a standard procedure, analogous to quantitative sensitivity analysis for meta-analysis. As such **Paper M3** contributes to an emerging evidence base and validates this explicit recommendation in Cochrane Supplementary Guidance (Hannes, 2011). Finally the observation that study quality may differ across different disciplines or for different sub-populations covered by a review (as exemplified by **Case Studies 2-4**) relates to issues of sampling (first raised in **Paper M1**) and to mechanisms of exploring heterogeneity (**Paper M5**).

4.2 Limitations of Paper M3

Paper M3 shares limitations reported by Harden (2008) in only using two systematic reviews opportunistically as case studies. However, the intention was to augment and to illuminate the evidence base, not to provide a step-wise methodological advance. Even though the explanatory power of two case studies is limited, previous explorations of this issue (Noyes & Popay, 2007; Thomas & Harden, 2008) reported a single case study. In addition, **Paper M3** provides an overtly methodological investigation of an issue only observed *in passim* within the two previous review reports. Harden's Methods Festival presentation (2008), identified while preparing this Thesis, does include experiences from five reviews but this data has not yet appeared in a peer-reviewed publication.

Paper M3 is subject to the limitations and biases associated with a retrospective investigation. Primarily the reason for a retrospective approach is that the practical constraints imposed by delivering reviews to a pre-specified deadline impair the ability of any review team to perform such a methodological analysis prospectively. Dixon-Woods, Bonas...Booth et al (2006) draw a contrast with quantitative syntheses before pointing to an associated logical limitation:

precisely how a sensitivity analysis for an interpretive synthesis could be undertaken is unclear. Once a paper has made its contribution to the development of concepts and theories, it may be difficult to simply extract it to see what the synthesis would look like without that paper.

A 'Dual Heritage' for QES

To a certain degree **Paper M3** side-steps such complexity by utilising two examples of thematic synthesis, a QES methodology positioned towards the aggregative end of a configurative-aggregative continuum (Gough et al, 2012a; 2012b). For this purpose a simplistic analysis of each paper's contribution to the final synthesis; namely "vote-counting" using a binary yes/no judgement, was judged sufficient. Interestingly, Harden reports results for her five included reviews on a quality continuum. Subsequently she reduces this more sophisticated analysis to a two-by-two matrix mapping low-high quality and low-high contribution. It is noteworthy that, as in **Paper M3**, Thomas and Harden performed sensitivity analysis procedures on *thematic* syntheses. The implication is that sensitivity analysis would be more complex within a more configurative method of QES, such as meta-ethnography. In the one published example that considers the sensitivity of review findings in a meta-ethnography, Atkins and colleagues (2008) focus on the early generation of themes, concluding:

study quality had an effect on the contribution of a paper to the overall synthesis. Papers that provided mainly descriptive data offered few insights, while others that included thick description and rigorous analysis contributed more substantively to the themes.

The implication is that subsequent second- and third-order constructs are similarly unaffected by a focus on studies with thick descriptions. However this should be tested empirically, particularly given that it gets progressively more difficult to establish an audit trail for synthetic constructs. Garside and colleagues (2008) offer an additional understanding of sensitivity analysis by considering the impact of a search update on an extant synthesis, concluding that, as a consequence, they could demonstrate saturation for their meta-ethnography, at least in terms of themes. However the conceptual richness of one of the subsequent papers would have enhanced the explanatory power of their synthesis. This confirms that for a "configurative" review, such as a meta-ethnography, it is the characteristics of an individual paper, not its existence *per se*, that determines its potential usefulness for the interpretation (Child et al, 2012).

Ostensibly, sensitivity analysis seeks to provide "objective information on the impact of methodologically sound studies versus studies that contain methodological flaws" (Hannes et al, 2010). However subjectivity is unavoidably introduced at several stages

A 'Dual Heritage' for QES

such as in judgements on how quality is to be assessed, which criteria are to be privileged and which threshold is to be implemented.

4.3 Summary

Quality assessment is a methodological “tectonic pressure point” where the dual heritages of systematic review and primary qualitative research come into shuddering juxtaposition. Epistemological and practical differences can be detected at every level of the debate, from what is meant by “quality” through the role of checklist and criteria to the appropriate course of action when studies fall short of minimal quality. **Paper M3** has demonstrated that, while exclusion of poorer quality studies may *generally* have minimal impact on the resultant synthesis, the review team must be sensitive to *particular* groups of studies, whether by discipline or perspective, that may be discriminated against by application of a quality threshold. **Paper M3** affirms that review teams should consider the appropriateness of testing the overall synthesis, through qualitative procedures analogous to subgroup analysis or sensitivity analysis, in order to challenge, and thus ultimately preserve, the integrity of the synthetic findings.

5. On Synthesis

Paper M2. Synthesizing qualitative research: a review of published reports

Paper M4. A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents.

Paper M5. Desperately Seeking Dissonance: Identifying the "Disconfirming Case" in Qualitative Evidence Synthesis.

Booth (2011b) draws upon earlier work by Mays et al (2005) in describing synthesis as "the stage of a review in which evidence extracted from different sources is *juxtaposed* to identify patterns and direction in the findings, or *integrated* to produce an overarching, new explanation or theory which attempts to account for the range of findings". This distinction between aggregative and interpretive, also referred to as configurative (Gough et al, 2012a; 2012b), functions provides a backdrop for examination of the 42 published QES in health and healthcare included in the dataset for **Paper M2**. **Paper M2** identified meta-ethnography, an essentially interpretive method, as the most prevalent form (n=19) of QES – posing a particular challenge given that the method does not claim a systematic review heritage. Furthermore **Paper M2** observed that many papers were mislabelled, creating further methodological confusion, a finding subsequently reported by other commentators (Paterson, 2011). **Paper M2** identified a need for models embodying a more 'organic, creative and interpretive approach to conducting reviews of complex literature'. In other words there was an imperative to use methods drawn from the primary qualitative heritage in tackling methodological issues not accommodated within the conventional systematic review template.

The ESRC End of Project Report highlighted two broad directions for further research identifying:

a real need to interrogate the epistemological assumptions underlying the inclusion of qualitative research in systematic reviews. There is an uneasy fit between the frame offered by conventional systematic review methodology and...epistemological assumptions and research practices associated with qualitative research (Dixon-Woods, Booth et al, 2007a).

As well as,

a need to...establish a set of principles and processes that might inform interpretive syntheses, as distinct from the kinds of aggregative syntheses that systematic review methodology has traditionally produced...(Dixon-Woods, Booth et al, 2007a)

These challenges set a backdrop for the remainder of the research presented in this Thesis.

5.1 Towards a pragmatic method of synthesis

Recent years have witnessed increasing acknowledgement of the potential contribution of QES to health technology assessment, signalled most noticeably by a methodological review from NHS Quality Improvement Scotland (Ring et al, 2010). Health technology assessments face particular constraints when seeking to optimise the trade-off between rigour and relevance (Rotstein & Laupacis, 2004), with relevance particularly evidenced in timeliness. They also face an additional imperative to produce a final product and outcome that is fundamentally pragmatic. The Author encountered such drivers in a project funded by the NHS Health Technology Assessment Programme which sought to enhance a meta-analysis of chemopreventive agents (Carroll, Cooper... Booth et al, 2010; Cooper, Squires... Booth et al, 2010) with qualitative evidence focusing on the views of adults taking such agents. The qualitative review sub-team (Carroll, Booth, and Cooper) responded by harnessing the pragmatic utility of framework synthesis. This involved using a conceptual model as a starting point for coding evidence from twenty included studies. **Paper M4** constitutes a description, analysis of the value, and expansion of this approach, labelled by the Author as "a best fit approach", in this chemoprevention case study.

The choice of framework synthesis (**Case Study 3**) was suggested by prior involvement as mentor for a qualitative systematic review (Lloyd-Jones, 2004; Lloyd-Jones, 2005). The utility of this method of synthesis corroborates the previously advanced opinion that many potential methods for qualitative synthesis derive from a primary qualitative research heritage (Booth, 2001). Framework analysis was developed by two qualitative researchers (Ritchie & Spencer, 1994). Its claimed

A 'Dual Heritage' for QES

advantages are consonant with the objectives of health technology assessment: “it is better adapted to research that has specific questions, a limited time frame, a pre-designed sample (e.g. professional participants) and a priori issues (e.g. organizational and integration issues) that need to be dealt with” (Srivastava & Thomson, 2009). All these characteristics led the sub-team to believe that the corresponding variant for secondary research, framework synthesis, would prove valuable in the context of a technology assessment.

The Author systematically searched for and identified a conceptual model, selected for its broad applicability to the review topic, as a starting point for an inductive process of analysis. New or adjusted concepts, not previously defined or explained by the model, were added as they were identified. The revised conceptual model represented a “best fit” model that “short circuited” the typically “lengthy process of model specification” (Dixon-Woods, 2011). This relative advantage, together with the contingent nature of the method, was spelt out explicitly when amending the final manuscript.

5.2 Exploring heterogeneity

The distinction between aggregative and interpretive (or configurative) approaches to systematic reviews is important (Dixon-Woods, Bonas...Booth, et al, 2006), even though such a polarization masks the fact that many syntheses use elements of both. Few have commented, however, that these approaches may be differentiated across a further dimension i.e. that aggregative approaches focus on identifying the *commonalities* within a body of studies (a key element of the systematic review heritage) while interpretive approaches, as characterized by the primary qualitative heritage, offer an opportunity to explore *differences*. This observation on aggregative approaches is true regardless of whether the final synthesis product is a formal meta-analysis or a narrative synthesis reporting the “epidemiology” of a set of studies. Essentially aggregative approaches seek to establish whether an intervention works “on average”. However the notional “average” is a statistical artifact that may represent a case (as in 2.2 children) that may not even exist.

A 'Dual Heritage' for QES

Experience from multiple reviews, for a variety of funding agencies, identified a need to develop systematic procedures to identify heterogeneity within qualitative syntheses. Such a need was stimulated by the observation by Petticrew & Roberts (2006) that:

we select, evaluate, and remember information in a way that supports our individual preferences, we fail to look for evidence that disconfirms our pet hypotheses, and we cannot spot errors in our own reasoning (p. 130).

What was apparent, and what the Author had previously signaled in revisions to **Paper M3**, was a need for approaches that allow for identification of heterogeneity. **Paper M5** identified fourteen such mechanisms for exploring heterogeneity and, with the Author as the common denominator, harnessed three different review teams in illustrating their practical potential and use. These mechanisms are grouped within four main categories (**Table 3**)

Table 3 - Strategies for exploring heterogeneity within a QES (from Paper M5)

Strategies relating to the review team:	Role of the review team, different reviewer backgrounds, reflexivity, team dynamics and legitimizing a culture of questioning.
Strategies relating to the review methods:	Use of methods for identification of studies, differential exposure to the findings, multiple readings, different methods of analysis, analytic procedures, qualitative sensitivity and subgroup analyses
Strategies relating to use of theory:	Testing of existing theories/frameworks
Strategies relating to presentation of findings:	Methods for presentation of findings, stakeholder or respondent validation

5.3 Role of the Case Studies

The practical challenge, faced in **Case Study 1**, of extracting and analysing data efficiently, particularly for the comparatively poorly resourced qualitative component of a health technology assessment, stimulated the Author to suggest the “best fit” approach to synthesis. It is interesting to contrast the function of framework synthesis in **Case Study 1**, where the review team acknowledged *a priori* the limited

A 'Dual Heritage' for QES

and conditional utility of the framework, with subsequent experience extracting data to a framework in **Case Study 5**. In the latter case, a framework based on a modified Federal Emergency Management Agency (FEMA) Emergency Planning cycle (McLoughlin, 1985), specifically drawn from the review context of emergency planning, could be considered comprehensive. The primary function of a framework was to highlight thematic areas in which there were research gaps or where a thematic area was relatively well-researched. In **Case Study 1** the focus of interest was on conceptual development and the initial framework became a “scaffold” upon which to hang thematic ideas, some already accommodated and others to be incorporated. In contrast, in **Case Study 5**, the framework served as a “window” through which the panorama of the topic could be viewed and apportioned. Both applications, drawing on the shared etymology of “framework”, are accommodated by framework synthesis or analysis. Finally **Case Study 3** employed inductive and deductive approaches, with one investigator pursuing each approach (**Paper M5**). This “manufactured experiment” helped to reveal mismatches between a policy framework for protocol based care (NHS Modernisation Agency, 2002) and the reality “on the ground”, reminding the reviewer that selection of pre-existing frameworks is a value-laden, not neutral, process. It also provided initial inspiration for the “best fit” approach which formally combines both inductive and deductive approaches.

A 'Dual Heritage' for QES

Table 4 - Characteristics of the Case Studies

Topic	Funder	Role of Author	Method of Synthesis	Publications (Citations)
Chemoprevention of colorectal cancer	National Institute for Health Research Health Technology Assessment Programme; 2007-2009	Co-Investigator, Methodologist and Third Reviewer	Best Fit Synthesis	Carroll et al, 2010a* (21 GS), 2010b (22 GS); Cooper et al, 2010* (31 WoS; 54 GS); Papaioannou et al, 2011 (6 GS); Squires et al, 2011 (2 GS).
Case Study 2 –Student experience of workplace-based e-learning	Higher Education Academy; 2007	Principal Investigator, Methodologist and Third Reviewer	Thematic Synthesis	Booth et al, 2009* (4 WoS; 14 GS); Carroll et al, 2009a* (20 GS), 2009b* (4 GS), 2011 (0 GS); Papaioannou et al, 2010* (14 GS); Booth, 2011d* (0 GS)
Case Study 3 – The nurse, midwife and health visitor contribution to protocol based care	National Institute for Health Research Service Delivery and Organisation Programme; 2006-2008	Methodologist and Second Reviewer	Thematic Synthesis and Framework Synthesis (in parallel)	Patterson et al, 2010 (0 GS); Ilott et al, 2006 (19 GS), 2010* (3 WoS; 6 GS)
Case Study 4 – Group therapies for postnatal depression)	National Institute for Health Research Health Technology Assessment Programme; 2007-2009	Co-Investigator, Methodologist and Second Reviewer	Thematic Synthesis	Stevenson et al, 2010a* (6 GS); 2010b (3 GS); Scope et al, 2012* (0 WoS; 0 GS)
Case Study 5 – The evidence base for emergency planning	National Institute for Health Research Service Delivery and Organisation Programme; 2010-2012	Co-Investigator, Methodologist and Reviewer	Framework Synthesis and Narrative Synthesis	Challen et al, 2012* (0 WoS; 0 GS); Lee et al, 2012a* (0 WoS; 0 GS); 2012b (0 WoS; 0 GS).

Key: * = Publications involving Author; WoS = Web of Science; GS = Google Scholar

A 'Dual Heritage' for QES

Experience acquired during three of the case studies (**Case Studies 1, 2 & 3**) directly informed thinking on exploring the deviant case (**Paper M5**) for improving the quality of subsequent analysis. Work undertaken to inform **Case Study 3** was the catalyst for this exploration, having discovered that the value of different synthesis methods was further augmented by bringing in a third, more-distanced observer. This dynamic was replicated for both **Case Study 1** and **Case Study 2**, adding value to the analysis by bringing in additional theoretical resources (**Case Study 1**) and by identifying tensions between themes (**Case Study 2**). While conventional systematic review methods identify roles for a second data extractor (to ensure consistent data extraction) and for a third, more-neutral observer (as an arbiter in cases of disagreement) these roles are markedly different within QES. In QES the second data extractor may provide a different interpretation, offering divergence rather than convergence. Similarly a *third* observer may offer a more detached interpretation that belongs to neither reviewer, justifiably leading to the synthesis being considered "more than the sum of its parts" (Barnett-Page & Thomas, 2009). Reinterpretation of reviewer roles within a qualitative paradigm may offer a "transformative" variant of the dual heritage.

5.4 Limitations of Papers M4 and M5

It is important to emphasize that the "best fit" approach represents a specific response to the short timeframe that characterizes policy development (Sweet & Moynihan, 2007). The limitations of the "best fit" method are common to those of framework analysis, its primary research precursor: "it is important to maintain an open mind and not force the data to fit the a priori issues" (Srivastava & Thomson, 2009). If time and circumstances allow then it is preferable to spend more time in sensitizing the research team to the topic and the literature, on specifying the theoretical framework, and in refining the coding, charting and mapping (Dixon-Woods, 2011). Preferably, too, time can be devoted to those elements of "creativity" embodied in the Analysis component of the SALSA mnemonic (Booth, Papaioannou et al, 2011). Notwithstanding these reservations Dixon-Woods has observed, in a methodological commentary on **Paper M4**, that:

A 'Dual Heritage' for QES

framework-based synthesis using the 'best fit' strategy is, in the right hands, likely to be a highly pragmatic and useful approach for a range of policy urgent questions (Dixon-Woods, 2011).

Such a formative verdict is being confirmed by emerging studies. For example a Protocol for Plain Tobacco Packaging: a Systematic Review (Moodie et al, 2011) proposes use of framework synthesis, citing **Paper M4** alongside work by Oliver et al (2008):

This process entails refining the framework, and confirming or refuting parts of it, as well as extending the framework and specifying the particular populations those specific parts of the framework does, or does not apply to. (Moodie et al, 2011)

Similarly a variant of “best fit” synthesis which “used the Health Belief Model (HBM) as a framework for analysis” for two Public Health Reviews for the National Institute for Clinical Excellence, further attests to the pragmatic value of framework synthesis within a time-constrained context (Lorenc et al, 2012). The ongoing development of framework-based approaches, as used by such Centres as the EPPI-Centre, University of London (Oliver et al, 2004; Kiwanuka et al, 2011) and the School of Health and Related Research (SchARR), University of Sheffield (e.g. **Case Study 1**) will likely further enhance their utility within a policy-urgent environment (Dixon-Woods et al, 2011; Bosch-Capblanch et al, 2012; Dixon-Woods et al, 2012a, 2012b; Lewin et al, 2012). “Best fit” synthesis may also offer a potential solution to an issue highlighted by Garside (2008), namely that identification of a suitable framework may, in practice, occur later than at the scoping phase as previously anticipated by Paterson and colleagues (2001). By acknowledging the provisional status of the framework, and by initiating procedures for identifying such a framework that are independent from sifting for the review itself, it becomes feasible to use a framework at an earlier juncture in the review process.

In addition to the acknowledged methodological limitations of the “best fit” approach, **Paper M4** was itself subject to important constraints. Experiences from a single review are described and it is not clear how feasible identification of a “best fit” model is for other topics. However the method offers more flexibility than specification of a purpose-specific model or framework and so possesses “relative

A 'Dual Heritage' for QES

advantage" (Rogers, 1995). The Author is currently exploring the challenge of *systematically* identifying theories or models (See *Chapter 8 -Towards a Future Research Agenda*).

Paper M5 similarly employs methods identified during the conduct of three pragmatic review projects, essentially case studies. It is worth flagging the irony of employing low level study designs from the evidence hierarchy in the quest to strengthen methods used to manufacture the evidence base, an irony compounded by the fact that case study designs, while overlooked by most systematic reviews, are considered a legitimate form of evidence within the qualitative paradigm. Nevertheless **Paper M5** is strengthened by a rapid, yet systematic, review of methodology as reported in published papers. While many authors comment on the need to examine heterogeneity within qualitative systematic reviews (Sandelowski et al, 2007), few have operationalized, or even suggested, procedures to achieve this. Arguably **Paper M5**, both for methodological innovation and academic leadership, represents the Author's principal contribution to date.

5.5 Summary

Synthesis is probably one of the most exciting areas of methodological innovation and seems to have most to gain from application of tools from the primary qualitative research heritage. Indeed in *Systematic Approaches to a Successful Literature Review*, Booth (2011b) lists a toolkit of tactics for testing or confirming review findings based on corresponding tactics (Miles & Huberman, 1994) identified for primary qualitative research. The Author particularly sought pragmatic methods for consistent and transparent data synthesis, evidenced in development of the 'best fit' approach (**Paper M4**), and the exploration of heterogeneity, evidenced in the quest for the disconfirming case (**Paper M5**). **Papers M4 and M5** occupy a place at the vanguard of current issues within QES.

6. Improving Standards of Reporting

Paper M1. "Brimful of STARLITE": toward standards for reporting literature searches.

Paper M2. Synthesizing qualitative research: a review of published reports.

Paper M3. Should we Exclude Inadequately Reported Studies from Qualitative Systematic Reviews? An Evaluation of Sensitivity Analyses in Two Case Study Reviews.

Systematic reviews have exerted a profound influence on standards and guidelines for reporting research. Such influence operates on the final published version of the review, with the requirement for clear, explicit and reproducible reporting of methods to increase confidence in the rigour of the review process. It further impacts by requiring that studies to be included in each systematic review must provide sufficient detail of their own methods and results to allow quality assessment and synthesis. Teams of reviewers have produced an array of guidelines prescribing the content and format of different types of study (Moher et al, 2011), rejoicing in quirky yet memorable acronyms (e.g. PRISMA, STARD, SQUIRE etcetera). In contrast, within the qualitative research community there is considerable antipathy to prescriptive guidelines for reporting qualitative inquiry. Qualitative reports are also seen to be disadvantaged due to constrictive word limits in many journals (Garside et al, 2010). This may make the provision of rich data and analysis, and indeed auditable methods, particularly problematic (Garside et al, 2009). Indeed several qualitative methods, including Interpretative phenomenological analysis (IPA) and "classical" grounded theory, are frequently criticised for insufficient details of their processes of data analysis. Nevertheless the requirements for transparent and explicit reporting signify that it may only be a matter of time before consensual standards are extended to reports of QES.

Several authors have identified particular issues with reporting of search procedures with several citing **Paper M1** to support the need for "greater detail for reporting... *sampling strategy*, *type of study*, *approaches*, *range of year*, *language*, *inclusions and exclusions*, *terms used*, and *electronic sources*" (Zhang et al, 2006) (**Table 4**).

A ‘Dual Heritage’ for QES

Table 5 - Elements of the STARLITE mnemonic (from Paper M1)

Element	Explanatory Notes
S ampling Strategy	<ul style="list-style-type: none"> • Comprehensive – attempts to identify all relevant studies on the topic • Selective – attempts to identify all relevant studies but only within specified limits • Purposive – samples from specific disciplines, years, journals
T ype of Studies	<ul style="list-style-type: none"> • Fully reported – describes actual study types (e.g. grounded theory) or designs to be included • Partially reported – uses an “umbrella” category such as “qualitative studies” without defining what this means.
A pproaches	<ul style="list-style-type: none"> • Approaches other than electronic subject searches (see below) • e.g. Handsearching • Citation snowballing
R ange of Years (Start Date-End Date)	<ul style="list-style-type: none"> • Fully reported – includes start and end date with justification for time period chosen • Partially reported – includes start and end date but determined only by available coverage of databases
L imits	<ul style="list-style-type: none"> • Functional limits that are applied for logistic reasons but do not alter the topic conceptually (e.g. Human, English etcetera)
I nclusion and Exclusions	<ul style="list-style-type: none"> • Conceptual limitations that mediate the scope of the topic area (e.g. geographical location, setting or a specific focus of study)
T erms Used	<ul style="list-style-type: none"> • Fully present: example of a sample search strategy from one or more of the main databases. • Partially present: reports terminology used but without evidence of search syntax and operators
E lectronic Sources	<ul style="list-style-type: none"> • Reports databases used and, optimally, search platforms and vendors to assist in replication

Deviation from a default of “comprehensive searching”, particularly involving greater reliance on iterative searching and supplementary approaches, makes it even more imperative to document both the “search strategy and, where used, the methods for sampling” (NHS Centre for Dissemination, 2008). The need for explicit reporting, as endorsed by **Paper M1**, has been restated recently (Finfgeld-Connett & Johnson, 2012). The Author is cautious in advancing the STARLITE structure as a standard, recognising that it is based on current, not consensual, best practice (Moher et al, 2011). However some QES authors have adopted STARLITE as a *de facto* standard (Edwards et al, 2010). Other reviewers have used the structure to report searches for quantitative systematic reviews (Cooperstein & Lew, 2009). One such report reads:

The methods used for conducting and reporting the literature search followed the approach proposed by the STARLITE investigators (Rambout et al, 2007)

A 'Dual Heritage' for QES

While another observes:

STARLITE was developed for qualitative studies but could also be useful when searching for quantitative studies (Burns & Chung, 2010)

The *Cochrane Handbook* cites **Paper M1** to support an “imperative to improve quality of reporting standards of search methods” (Noyes,...Booth et al, 2008). **Paper M2** has similarly been cited to support the need to further define the contents of search reports (Fehrmann & Thomas, 2011). Central to this issue are previously articulated concerns (Booth, 2001) as to whether quantitative systematic reviews provide an appropriate template for qualitative evidence syntheses:

it is at present difficult to distinguish whether this is because of genuine differences of methodological principle between systematic review and synthesizing qualitative research (which are likely to be defensible) or differences in explicitness and quality of execution (which are likely to be less defensible) [**Paper M2**].

We subsequently articulated this distinction between what is defensible and what is indefensible:

it may well be preferable for syntheses of qualitative research to use methods of searching, appraisal, and synthesis that are quite distinct from those used in (conventional) systematic review; what is less acceptable is when these methods are not adequately justified or are poorly undertaken or reported [**Paper M2**]

However, movement towards better standards of reporting may, in turn, stimulate methodological development. Indeed the writing team for **Paper M2** debated the “chicken and egg” nature of such progress, with the Author’s view that the field was still in a descriptive (rather than prescriptive) phase eventually prevailing in the final manuscript:

A move towards improved explicitness about reporting of syntheses of qualitative research could take place ahead of a consensus emerging on methods for synthesis, and would likely contribute to improved reflexivity and better research practice [**Paper M2**]

Subsequently, the Author has taken forward the imperative for improved reporting of primary qualitative papers in a chapter co-written with the convenors of the Cochrane Qualitative Methods Group in *Reporting Health Research* (Booth et al, forthcoming) for the EQUATOR Network. This showcases and provides a critical

commentary on the 32-item **Consolidated criteria for reporting qualitative research (COREQ)** checklist (Tong, 2007). Subsequently the originator of the checklist for primary qualitative research has initiated an **Enhancing Transparency in REporting the synthesis of Qualitative research (ENTREQ)** statement, a draft standard for reporting qualitative syntheses (Tong et al, 2012a; 2012b), to be further evaluated by the Cochrane Qualitative Methods Group.

6.1 Summary

While concerns regarding the reporting of primary qualitative research have understandably taken a secondary place to methodological development they underpin several areas investigated in this portfolio. Study identification requires adequate listing of data sources and delineation of keywords, search syntax and strategies (**Paper M1**). Quality assessment of primary qualitative studies enables a review team to form judgements on the robustness of the evidence base (**Paper M3**). It is tempting to locate qualitative research more towards the “knowledge support” end of the continuum and argue for its manifest lack of utility in a pragmatic context of “decision support” (Mays et al, 2005). In truth, however, as with published reports of all types of research study, individual studies may be located anywhere on this continuum. The utility of a research study is not determined either by the pragmatic nature of its design or methods or by the clarity and detail of its reporting. However deficiencies in design, methods or reporting may impair a study's overall usefulness.

Meanwhile any methodological survey of published syntheses (**Paper M2**) necessarily depends on the quality of published reports. Hannes & Macaitis (2012) recreated the methods of **Paper M2** (for the period 2005 to 2008) to confirm:

a black box between what people claim to use as a synthesis approach and what is actually done in practice.... the boundaries between methods of synthesis have become blurred in scientific literature (Hannes & Macaitis, 2012)

Once reporting standards are introduced and accepted within their research community, they provide an opportunity to audit compliance and document variation, as for quantitative systematic reviews (e.g. Hind & Booth, 2007). This may

A 'Dual Heritage' for QES

stimulate further exchange of methods and, as a consequence, may contribute to methodological clarity and innovation.

7. Some Reflections on the Dual Heritage model

As already made clear, this portfolio represents an opportunistic and pragmatic response to particular challenges faced in the course of delivering funded reviews. Nevertheless reflection on lessons from these included papers has enabled the Author to identify four different “models” by which the dual heritage of qualitative and systematic review research methods might interact. While simply outlining the broad characteristics of these models risks the danger of caricature, they do help to indicate the versatility of approaches on which to draw. Broadly speaking, as with other variants of mixed methods research, the two heritages can combine to expand the utility of the synthesis product and thus offset the weaknesses of either approach in isolation (Driscoll et al, 2007).

In some cases the heritages present genuine methodological choices (an **alternatives model**). For example, a review team may minimise the impact of context, stripping all bar the basic epidemiological details relating to Setting and extracting these into tables, in a manner analogous to the work that precedes meta-analysis. Alternatively contextual variation may constitute an essential part of the interpretation, in which case context needs to be explored and individual studies situated. Similarly the review question may be fixed and prespecified, as with the **P**opulation, **I**ntervention, **C**omparison and **O**utcome (**PICO**) formulation of systematic reviews (Sayers, 2008). This assumption certainly underpinned many early QES and led the author (Booth, 2004, 2006) to develop a corresponding **S**etting **P**erspective phenomenon of **I**nterest **C**omparison and **E**valuation (**SPICE**) formulation, adopted for QES and methodology reviews (Malpass et al, 2009; Lewis et al, 2010; Jones et al, 2011, Windle et al, 2011). Alternatively, it may be equally valid for the review question to emerge iteratively from the data as with primary grounded theory approaches (Barnett-Page & Thomas, 2008). In parallel there is an emerging menagerie of attempts to develop and extend question formulation approaches (Davies, 2011), for example PICOC (Petticrew & Roberts, 2006) and SPIDER (Cooke, Smith & Booth, 2012).

A 'Dual Heritage' for QES

In other cases the two heritages may come into play at different stages of the review process (a **sequential model**); for example, a QES may start by comprehensively sampling the literature, as per the conventional systematic review model, to construct an overall sampling frame. Subsequently the sampling strategy may employ purposive or theoretical sampling approaches from qualitative research, in order to explore particular interpretations or productive lines of inquiry.

A third circumstance is where a tool or technique developed within one heritage, for example sensitivity analysis, is “translated” or interpreted appropriately within a new methodological context (a **transformative model**). The intention is not to replicate the source “method” but rather to acknowledge a shared concern by developing an analogous counterpart. The challenge in such transformation is to satisfy the rigour and transparency required by the systematic review heritage but in a way that is sensitive to the heritage of primary qualitative research.

Finally the two heritages may work together, with each contributing meaningfully to an end product that is greater than the sum of its parts (a **synergistic model**). For example, the PRISMA standards of reporting (from the systematic review heritage) (Liberati et al, 2009) may contribute auditability while methods of presenting thematic analysis (from the qualitative research heritage) may enrich interpretation of the synthesis product (Pope et al, 2007). Working side by side the two heritages manufacture a refined product that draws from each tradition. Clearly the challenge is not, as first supposed, to select one of these models as a dominant influence on QES methods. Instead the richness of both heritages is best exploited in choosing from such models judiciously, whether for one or more stages of the review process or for a review in its entirety.

8. Towards a Future Research Agenda

The foregoing commentary charts personal and collective progress in development of QES methods. However substantive methodological issues remain to be explored. These constitute a future research agenda, to be carried forward, by the Author and colleagues:

I. Empirical work on the implications of sampling alternatives to comprehensive searching for the rigour of the resultant synthesis.

With an ever wider range of qualitative synthesis methods, many advocating purposive or theoretical sampling methods, there is a need to develop rigorous procedures for constructing an appropriate sampling frame. Subsequently there is a need to test the interpretive value of the resultant synthesis product in comparison to a similar review that includes a comprehensive and exhaustive sample of studies.

II. Investigation of supplementary search techniques to complement protocol-driven searches for qualitative syntheses.

Having explored the potential value of supplementary search techniques, within a specific qualitative review of e-learning (Papaioannou,...Booth et al, 2010), there is scope for further exploration. With 70-85% of included studies from a wide range of review topics, including those involving qualitative research, available via a single database search of MEDLINE (Booth, 2012. Unpublished data) the Author intends to extend the investigation, commenced in **Paper M1**, to focus on how to identify most efficiently the remaining 15-30% of studies, particularly with a view to diversifying the sample. Candidate methods include those employing 'sibling searching' (Booth, 2011a), e.g. seeking qualitative reports from named programmes, or from references listed in studies, previously identified for an effectiveness review (Garside et al, 2009).

III. Prospective investigation of the differential effect of primary study quality on the robustness of qualitative syntheses.

Following the findings of **Paper M3**, which extends case study-based work by other investigators, there is a need to accumulate further evidence informing decisions

around the quality of included studies. Such an investigation will likely inform an understanding of what exactly study quality means, for example in relation to conceptual richness or “thickness of description”, within qualitative research.

IV. Specific methods and strategies for the systematic identification of theories and models as the basis for framework or best fit synthesis.

Within healthcare current approaches to identification of models and theories for use within a systematic review appear serendipitous and casual in comparison to the specified and auditable stages associated with study identification. Additionally candidate theories accessed by a particular review project seem overly reliant on the intellectual resources and background of team members. Having demonstrated (**Paper M4**) that extraction of data against a pre-specified framework (whether pre-existing as a framework or, alternatively, deconstituted – and then ultimately reconstituted - from a model or theory) is both efficient and flexible, even when such a framework is only contingent and “best fit”, reviewers need to identify candidate theories and models in a systematic manner (Lorenc et al, 2012).

V. Exploration of the relative value of different methods for identification of heterogeneity.

Although awareness of the importance of issues relating to study heterogeneity within the context of QES is widespread, few commentators specify, or empirically investigate, methods for exploring such heterogeneity. It is timely to investigate the comparative usefulness of the more feasible or practicable candidate methods, as suggested by **Paper M5** albeit with few instances of their actual use or documentation.

VI. Empirical contribution to the evaluation of reporting standards for qualitative syntheses and for included individual studies.

Publishing standards, such as CONSORT and PRISMA (formerly QUOROM) (Liberati et al, 2009) have had a discernible impact upon the methodological quality, of different types of research output. It is opportune to evaluate appropriate guidelines for primary qualitative research and for qualitative syntheses and, then, to repeat

A 'Dual Heritage' for QES

methodological surveys such as **Paper M2** (and that by Hannes & Macaitis, 2011) to monitor the effect of such standards.

While these issues are important for QES in general, two emerging contexts provide a specific backdrop to future research. First, increasing interest in the evaluation of complex interventions requires the development of more flexible, iterative and creative approaches to the exploration and integration of issues identified from the qualitative evidence base (Shepperd et al, 2009). Second, increasing time and resource pressures are shaping an ever expanding range of review products (for example, the rapid evidence assessment (Ganann et al, 2010), the rapid health technology assessment (Booth et al, 2011) and the evidence summary (Khangura et al, 2012). These, in turn, require the development of methods of synthesis that optimise rigour and relevance (Laupacis & Straus, 2007) and that evaluate the consequences of pragmatic methodological choices (e.g. Watt et al, 2008). Lorenc and colleagues (2012) identify the potential benefits of such a confluence:

SRs could draw on the insights of other evidence synthesis methodologies regarding the benefits of iterative and theory-led approaches, whereby emerging constructs inform the selection of data for future stages of the review.

These imperatives of increasing complexity, and time and resource pressures, combine within the fundamentally pragmatic focus of the research presented in this Thesis. Indeed the “mixed heritage model” bears many hallmarks of the pragmatic school of thought, which maintains that “a false dichotomy exists between qualitative and quantitative approaches and that researchers should make the most efficient use of both [approaches] in understanding social phenomena” (Creswell, 1994). Analogous pragmatism, in reconciling systematic reviews and primary qualitative research, pervades this Thesis from preference for “qualitative evidence synthesis” over the purist “qualitative research synthesis” (Major & Savin-Baden, 2010) through to the deliberate mixing of deductive and inductive approaches within the “best fit method”. Such pragmatism also sidesteps, while not being oblivious to, important concerns regarding the desirability, or otherwise, of combining qualitative studies from different points on a descriptive-interpretive continuum.

9. Conclusion

This portfolio advances the Author's contribution to development of QES methodology. From initial interest in all stages of the QES process, signaled by the *Cochrane or Cockeyed* conference paper (Booth, 2001), the portfolio has pursued two hypotheses:

- (i) That qualitative evidence syntheses have much to gain from drawing upon the traditions and methods of primary qualitative research in tackling and overcoming practical methodological challenges
- (ii) That once a 'dual heritage model' is legitimized, through literature review and empirical methodological research, the path becomes clear to challenge key assumptions from the conventional systematic review 'template', leading to further methodological innovation.

Taken together, **Chapters One to Eight** provide an overview of how the dual heritage of QES has evolved. This dual heritage model is pervasive impacting, *inter alia*, on whether the review question should be fixed or negotiable (Eakin & Mykhalovskiy, 2003), the iterative nature of searching (Brunton et al, 2012; Finfgeld-Connett & Johnson, 2012), and the presentation of results (Harden et al, 2004) (**Table 1**). This Thesis focuses on four critical stages; namely study identification, quality assessment, synthesis and presentation. This requires a perspective, beyond that traditionally held by an information specialist, towards understanding the underlying rationales for different sampling approaches (Brunton et al, 2012; Suri, 2011).

Chapter Three identified limitations of current QES methods specific to literature searching and for the review process more generally. **Paper M1** challenged the notion that a comprehensive search strategy will always be appropriate and asserted that reporting of a sampling strategy is prerequisite to judging the appropriateness of a review. **Paper M2** identified methodological tensions as published QES lay claim to particular synthesis methods to establish their credentials but, in fact, undermine such claims by a lack of clarity in conduct, reporting or indeed both.

Chapter Four illustrated how it is productive to work within a 'dual heritage' model in that a procedure found to be useful in the context of quantitative systematic reviews, namely sensitivity analysis, could be "transformed" to examine the differential impact of study quality for an interpretive, or more accurately, 'configurative' (Gough et al, 2012a; 2012b) review product. Meeting a need again identified from a literature survey (**Paper M2**), Paper M3 highlighted that conceptually 'thick' papers were more likely to make a contribution than their thinner counterparts. However **Paper M3** also refocused this debate to recognize that the differential reporting conventions of different disciplines, professions or literatures might lead removal of papers on quality grounds to impact unfavourably on the external validity of the review, particularly for specific subpopulations.

Chapter Five turned its attention to the "engine room" of the synthesis product, namely the process of synthesis itself. It illustrated how the pressures of time, resources and a narrow policy window place a requirement for review methods that optimize rigour and timeliness. (Barroso et al, 2006) Whereas quantitative systematic reviews respond to these pressures through "internal" structures (described by Pawson as "the quart-into-pint-pot task of presenting the mass of data into to an intelligible set of summary matrices and tables" (Pawson, 2006b)) the specific requirements of QES reveal the value of introducing an external framework as a mechanism for rapid synthesis and analysis (Oliver et al, 2012). Notwithstanding acknowledged methodological limitations, shared with framework analysis, the framework-based approach possesses relative advantage even when the initial framework is not ideal but represents a contingent best fit (**Paper M4**).

Chapter Five also probes more deeply in exploring a fundamental difference in the underlying direction of travel of aggregative and configurative reviews, namely in their respective quests for homogeneity or heterogeneity. Just as primary quantitative research seeks to control heterogeneity while primary qualitative research delights in exploring complexity, so too qualitative systematic reviews are suited to identifying and exploring heterogeneity, instead of simply testing for its

A 'Dual Heritage' for QES

presence. However to succeed in this aspiration a reviewer must engineer active mechanisms to identify heterogeneity into QES processes (**Paper M5**).

Chapter Six examines study reporting; an issue impacting at all levels of the review process. Poor reporting may lead to poor indexing and may impair retrieval of relevant qualitative reports. It may make quality assessment of primary studies more challenging. At a synthesis level it can lead to obfuscation of review methods and therefore impair transfer of methods and innovation. While there is circularity in the relationship between improved methods and improved reporting, experience from other reporting standards indicates that progress in reporting may stimulate the construction of shared methodological understanding.

Chapter Seven provides an opportunity to use learning from the disparate papers in this portfolio to reflect on how the two components of the 'dual heritage' might interact, namely, the **alternatives**, **sequential**, **transformative** and **synergistic** models. It concludes that no single model captures the variety with which both heritages can contribute to viable pragmatic QES methods.

Chapter Eight builds upon the acknowledged limitations of the Included Papers in this Thesis by identifying opportunities for further empirical research to address unresolved methodological issues. Thus, this Thesis may be considered as a cross-sectional stopping post on a route with yet far to travel.

Taken as a whole the accelerated progress of QES, as documented in this Thesis, provides a refreshing antidote to the paradigm wars formerly waged, and still in isolated outbreaks ongoing, within primary research. The initial challenges raised by the *Cochrane or Cockeyed* paper (Booth, 2001) were deliberately provocative and confrontational, and raised more questions than answers. With increasing acceptance of the complementarity of insights from patients, carers, service users and clinicians, as captured in qualitative research (Jones, 2004), to the dominant focus of effectiveness reviews, the impetus is now more constructive. Faced with a dual heritage, of conventional systematic review methods and primary qualitative

A 'Dual Heritage' for QES

research techniques, exponents of QES may select judiciously from competing techniques, adapt from the richness of both traditions or maintain an open dialogue around viable alternatives. This illustrates that “placing [quantitative/qualitative] approaches in opposition does a great disservice by detracting from the contribution to be made by each, including what each can contribute to the other” (Wolcott, 2001). The rapprochement of the two heritages is heralded by a recent case study (Lorenc et al, 2012) that recognises the unique contribution to be derived from each source. In welcoming this team’s conclusion, namely “that such reviews are, to some extent, methodologically *sui generis* and cannot be governed *solely* [Italics added] by concepts imported either from SRs of quantitative evidence (e.g. comprehensiveness) or from primary qualitative research (e.g. saturation)” (Lorenc et al, 2012), the Author looks forward to operating within a methodological state of flux that offers stimulus for years to come.

References

- Atkins, S, Lewin, S, Smith, H, Engel, M, Fretheim, A, & Volmink, J (2008) Conducting a metaethnography of qualitative literature: lessons learnt. *BMC Medical Research Methodology* 8:21.
- Bambra, C (2011) Real world reviews: a beginner's guide to undertaking systematic reviews of public health policy interventions. *Journal of Epidemiology & Community Health*, 65(1):14-9.
- Barbour, RS (2001) Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *BMJ*, 322(7294), 1115-7.
- Barnett-Page, E, & Thomas, J (2009) Methods for the synthesis of qualitative research: a critical review. *BMC Medical Research Methodology*, 9:59.
- Barroso, J, Sandelowski, M, & Voils CI (2006) Research results have expiration dates: ensuring timely systematic reviews. *Journal of Evaluation in Clinical Practice*, 12(4):454-62.
- Baxter, S, Killoran, A, Kelly, MP, & Goyder E (2010) Synthesizing diverse evidence: The use of primary qualitative data analysis methods and logic models in public health reviews. *Public Health*, 124: 99-106.
- Beverley, CA, **Booth, A**, & Bath, PA (2003) The role of the information specialist in the systematic review process: a health information case study. *Health Information & Libraries Journal* 20(2):65-74.
- Boeije HR, van Wesel F, & Alisic E (2011) Making a difference: towards a method for weighing the evidence in a qualitative synthesis. *Journal of Evaluation in Clinical Practice*, 17(4):657-63.
- Booth, A** (2001) *Cochrane or cock-eyed? How should we conduct systematic reviews of qualitative research?* In: Proceedings of the Qualitative Evidence-based Practice Conference, Taking a Critical Stance. Coventry University, May 14 2001.
- Booth, A.** (2004), Chapter 6 - Formulating answerable questions, in Booth, A. and Brice, A. (Eds), *Evidence Based Practice for Information Professionals*, Facet Publishing, London, 61-70.
- Booth, A.** (2006). Clear and present questions: formulating questions for evidence based practice. *Library Hi Tech*, 24, 355-368.
- ***Booth A (2006). "Brimful of STARLITE": toward standards for reporting literature searches. *Journal of the Medical Library Association* Oct;94(4):421-9**

A 'Dual Heritage' for QES

- Booth, A** (2007) Author's Reply to Sandelowski M. In response to "Brimful of STARLITE". *Journal of the Medical Library Association*, 95(3):233
- Booth A** (2010) How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. *International Journal of Technology Assessment in Health Care*. 26(4):431-5.
- Booth A** (2011a) Chapter 3: Searching for Studies. In: Noyes J, Booth A, Hannes K, Harden A, Harris J, Lewin S, & Lockwood C (editors), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group. Available from URL <http://cqrmg.cochrane.org/supplemental-handbook-guidance>
- Booth, A** (2011b) Chapter 8 – Synthesising Included Studies. In: Booth A, Papaioannou D & Sutton, A *Systematic Approaches to a Successful Literature Review*. London: Sage,
- Booth, A** (2011c) Chapter 9 – Writing up and presenting data In: Booth A, Papaioannou D & Sutton, A *Systematic Approaches to a Successful Literature Review*. London: Sage, 197-222.
- Booth, A** (2011d) Chapter 10 - Managing the literature project. In: Booth A, Papaioannou, D & Sutton A. *Systematic Approaches to a Successful Literature Review*. London: Sage: 223-240.
- ***Booth A, Carroll C, Ilott I, Low LL & Cooper K (2013) Desperately Seeking Dissonance: Identifying the “Disconfirming Case” in Qualitative Evidence Synthesis. *Qualitative Health Research* 23 (1): 126-141.**
- Booth A, Carroll C, Papaioannou D, Sutton A, & Wong R** (2009). Applying findings from a systematic review of workplace-based e-learning: implications for health information professionals. *Health Information & Libraries Journal*, 26(1):4-21.
- Booth, A, Harris J & Hannes K** (2011) *Searching, sampling and selecting for qualitative evidence synthesis*. Workshop WS67 at Cochrane Colloquium, Madrid.
- Booth, A, Papaioannou, D & Sutton, A** (2011) *Systematic Approaches to a Successful Literature Review*. London: Sage.
- Booth, A et al** (forthcoming) Reporting Qualitative Research In: Moher D et al (eds) *Reporting Health Research*. London: Wiley-Blackwell.
- Bosch-Capblanch, X, Lavis, JN, Lewin, S, Atun, R, Røttingen, JA, et al. (2012) Guidance for evidence-informed decisions about health systems: rationale for and challenges of

A 'Dual Heritage' for QES

- guidance development. *PLoS Medicine* 9: e1001185.
doi:10.1371/journal.pmed.1001185.
- Brunton G, Stansfield C & Thomas J (2012) Finding relevant studies. In Gough D, Oliver S, Thomas J. *An Introduction to Systematic Reviews*. London: Sage Publications, 107-135.
- Brunton, G, Wiggins, M & Oakley, A (2010) *Becoming a Mother: A research synthesis of women's views on the experience of first-time motherhood*. ESRC End of Award Report, RES-000-22-2250. Swindon: Economic and Social Research Council.
- Burns, PB & Chung, KC (2010) Developing good clinical questions and finding the best evidence to answer those questions. *Plastic & Reconstructive Surgery*, 126(2), 613-8.
- Candy, B, King, M, Jones, L, Oliver, S (2011) Using qualitative synthesis to explore heterogeneity of complex interventions. *BMC Medical Research Methodology*. 11, 124. Available at: <http://www.biomedcentral.com/1471-2288/11/124>
- *** Carroll C, Booth A, & Cooper K (2011) A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Medical Research Methodology*, Mar 16;11:29. Available at: <http://www.biomedcentral.com/1471-2288/11/29>**
- ***Carroll C, Booth A, Lloyd-Jones M (2012) Should we Exclude Inadequately-reported Studies from Qualitative Systematic Reviews? An Evaluation of Sensitivity Analyses in Two Case Study Reviews, *Qualitative Health Research* , 22 (10): 1425-1434.**
- Carroll C, **Booth A**, Papaioannou D, Sutton A, & Wong R (2009a) UK Healthcare professionals' experience of e-learning techniques: a systematic review of qualitative data. *Journal of Continuing Education in the Health Professions*, 29(4): 2135-2141.
- Carroll C, **Booth A**, Papaioannou D, Sutton A, & Wong R (2009b). Enhancing the experience of e-learning among working students: a systematic review with thematic analysis, *IMPACT: Journal of Applied Research in Workplace E-learning*, 1(1): 85-96.
- Carroll C, **Booth A**, Papaioannou D, Sutton A, & Wong R (2011) Experiences of e-learning and its delivery among learners who work: A systematic review, in: *Supporting Workplace Learning: Towards Evidence-based Practice* eds. Poell R & Van Woerkom M. Springer; New York: 47-67.

A 'Dual Heritage' for QES

- Carroll, C, Cooke, J, **Booth, A**, Beverley, C (2006) Bridging the gap: the development of knowledge briefings at the health and social care interface. *Health Soc Care Community*, 14(6), 491-8.
- Carroll, C, Cooper, K, Papaioannou, D, Hind, D, Tappenden, P, Pilgrim, H, & **Booth, A** (2010a) Meta-analysis: folic acid in the chemoprevention of colorectal adenomas and colorectal cancer. *Aliment Pharmacol Ther*, 31(7), 708-18.
- Carroll C, Cooper K, Papaioannou D, Hind D, Pilgrim H, & Tappenden P (2010b) Supplemental calcium in the chemoprevention of colorectal cancer: a systematic review and meta-analysis. *Clin Ther*, 32(5):789-803.
- Carroll C, Lloyd-Jones M, Cooke J, & Owen J (2012) Reasons for the use and non-use of school sexual health services: A systematic review of young people's views, *Journal of Public Health*, 34(3): 403-10.
- Centre for Reviews and Dissemination (CRD) (2008) *Systematic Reviews: CRD's Guidance for undertaking reviews in health care*. University of York: Centre for Reviews and Dissemination.
- Challen K, Lee AC, **Booth A**, Gardois P, Buckley Woods H, & Goodacre SW (2012) Where is the evidence for emergency planning: a scoping review. *BMC Public Health*, 12(1):542
- Chalmers, I. & Altman, D (1995) (eds) *Systematic Reviews*. London: BMJ Publishing Group.
- Chalmers, I, Hedges, LV, & Cooper, H. (2002) A brief history of research synthesis. *Evaluation and the Health Professions*, 25(1), 12–37.
- Chapple, A, & Rogers, A (1998) Explicit guidelines for qualitative research: a step in the right direction, a defence of the 'soft' option, or a form of sociological imperialism? *Fam Pract*, 15(6), 556-61.
- Child S, Goodwin V, Garside R, Jones-Hughes T, Boddy K, & Stein K (2012). Factors influencing the implementation of fall prevention programmes: a systematic review and synthesis of qualitative studies. *Implement Sci*. 7(1):91.
- Cook DJ, Mulrow CD, & Haynes B (1998). Chapter 1 -Synthesis of best evidence for clinical decisions – Table 1.1. Differences between Narrative reviews and Systematic Reviews (p.7). In: Mulrow CD, Cook DJ, editors. *Systematic reviews, synthesis of best evidence for health care decisions*. Philadelphia: American College of Physicians; 1998. p. 5–12.
- Cooke A, Smith D, & **Booth A** (2012) Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qualitative Health Research*. 22(10):1435-43.

A 'Dual Heritage' for QES

- Cooper H (2007). *Evaluating and Interpreting Research Syntheses in Adult Learning and Literacy*. National College Transition Network, New England Literacy Resource Center/World Education, Inc., Boston, Mass.
- Cooper, K, Squires, H, Carroll, C, Papaioannou, D, **Booth, A**, Logan, RF, Maguire, C, Hind D, & Tappenden, P (2010) Chemoprevention of colorectal cancer: systematic review and economic evaluation. *Health Technol Assess*, 14(32), 1-206.
- Cooperstein, R, & Lew, M (2009) The relationship between pelvic torsion and anatomical leg length inequality: a review of the literature. *Journal of Chiropractic Medicine*, 8(3), 107-18.
- Corbin-Staton, AP. (2009) *Contexts of Parental Involvement: An Interpretive Synthesis of Qualitative Literature Using the Meta-Interpretation Method*. Dissertation Submitted to Faculty of the Graduate School of Education and Human Development of The George Washington University in partial fulfillment of requirements for degree of Doctor of Education. Available at: <http://gradworks.umi.com/3344879.pdf>. Accessed on 18/07/2012
- Creswell, JW (1994). *Research Design: Qualitative and Quantitative Approaches*. London: Sage.
- Davies, K (2011). Formulating the evidence based practice question: a review of the frameworks. *Evidence Based Library and Information Practice*, 6, 75–80
- Deeks, JJ, Glanville, J, & Sheldon, TA (1996) *Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews*. CRD Report 4, NHS Centre for Reviews and Dissemination, University of York.
- Dixon-Woods, M (2011) Using framework-based synthesis for conducting reviews of qualitative studies. *BMC Medicine*, 9:39.
- Dixon-Woods, M, Bonas, S, **Booth, A**, Jones, DR, Miller, T, Sutton, AJ, Shaw, RL, Smith, JA, & Young, B (2006) How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research* 6, 27-44.
- Dixon-Woods, M, **Booth, A**, Jones, D, Miller, T, Shaw, R, Smith, J, Sutton, A, Young, B (2007a), *How can systematic reviews incorporate qualitative research?* ESRC End of Project Report. ESRC. Available from: <http://www.esrc.ac.uk/my.../4816fee1-b170-4406-a759-c129eacc2ed7> Accessed on: 18/07/2012.
- ***Dixon-Woods, M, Booth, A & Sutton AJ (2007). Synthesizing qualitative research: a review of published reports *Qualitative Research*, Aug 7: 375-422.**
- Dixon-Woods, M, Cavers, D, Agarwal, S, Annandale, E, Arthur, A, Harvey, J, Hsu, R, Katbamna, S, Olsen, R, Smith, L, Riley, R, & Sutton, AJ (2006) Conducting a critical

A 'Dual Heritage' for QES

- interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*, 6:35.
- Dixon-Woods, M, McNicol, S & Martin, G (2012a) Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature. *BMJ Quality & Safety*, Apr 28. [Epub ahead of print]
- Dixon-Woods, M, McNicol, S & Martin, G (2012b) *Overcoming challenges to improving quality: Lessons from the Health Foundation's improvement programme evaluations and relevant literature*. London: Health Foundation.
- Dixon-Woods, M, Shaw, RL, Agarwal, S, & Smith, JA (2004) The problem of appraising qualitative research. *Qual Saf Health Care*, 13(3), 223-5.
- Dixon-Woods, M, Sutton, A, Shaw, R, Miller, T, Smith, J, Young, B, Bonas, S, **Booth, A**, & Jones, D (2007) Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *Journal of Health Services Research & Policy* 12(1), 42-7.
- Doyle LH (2003) Synthesis through meta-ethnography: paradoxes, enhancements, and possibilities. *Qualitative Research* 3 (3): 321-344
- Driscoll DL, Appiah-Yeboah A, Salib P, & Rupert DJ (2007) Merging qualitative and quantitative data in mixed methods research: How to and why not. *Ecological and Environmental Anthropology* 3:19–28.
- Eakin, JM. & Mykhalovskiy, E (2003) Reframing the evaluation of qualitative health research: reflections on a review of appraisal guidelines in the health sciences. *Journal of Evaluation in Clinical Practice* 9, 187–194.
- Edwards, A, Pang, N, Shiu, V, & Chan, C (2010) The understanding of spirituality and the potential role of spiritual care in end-of-life and palliative care: a meta-study of qualitative research. *Palliative Medicine*, 24(8), 753-70.
- Fehrmann P & Thomas J (2011) Comprehensive computer searches and reporting in systematic reviews. *Research Synthesis Methods*, 2 (1), 15–32.
- Fingeld-Connett D (2008) Meta-synthesis of caring in nursing. *Journal of Clinical Nursing*, 17, 196-204.
- Fingeld-Connett D & Johnson ED (2013) Literature search strategies for conducting knowledge-building and theory-generating qualitative systematic reviews. *Journal of Advanced Nursing*, 69(1):194-204.
- Ganann, R, Ciliska, D & Thomas, H (2010). Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science*, 5, 56.

A 'Dual Heritage' for QES

- Garside, R (2008). *A comparison of methods for the systematic review of qualitative research: two examples using meta-ethnography and meta-study* [PhD]. Exeter: Peninsula Postgraduate Health Institute, Universities of Exeter and Plymouth.
- Garside, R, Britten, N, & Stein, K (2008). The experience of heavy menstrual bleeding: a systematic review and meta-ethnography of qualitative studies. *Journal of Advanced Nursing* 63(6):550-62.
- Garside R, Pearson M, Moxham T, & Anderson R (2009) *Barriers to, and facilitators for, the effectiveness of multiple risk factor programmes aimed at reducing cardiovascular disease within a given population: a systematic review of qualitative research*. National Institute for Health and Clinical Excellence, London.
- Garside R, Pearson M, & Moxham T (2010) What influences the uptake of information to prevent skin cancer? A systematic review and synthesis of qualitative research. *Health Education Research*, 25(1):162-82.
- Glanville, J (1994) Evidence-based practice: the role of the NHS Centre for Reviews and Dissemination. *Health Libraries Review*, 11(4), 243-51.
- Gough, D, Oliver, S & Thomas, J (2012a) Introducing Systematic Reviews in Gough, D, Oliver S & Thomas J. *Introduction to Systematic Reviews*. London: Sage.
- Gough D, Thomas J, & Oliver S. (2012b) Clarifying Differences between Review Designs and methods, *Systematic Reviews*, 1 (1), Paper 28. Available at: <http://www.systematicreviewsjournal.com/content/1/1/28>
- Grant MJ, & Booth, A (2009) A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal* 26(2), 91-108.
- Greenhalgh, T, & Peacock, R (2005) Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*, 331(7524), 1064-5.
- Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O, & Peacock R (2005). Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Social Science & Medicine*, 61(2):417-30.
- Hannes, K (2011) Chapter 4: Critical appraisal of qualitative research. In: Noyes J, Booth A, Hannes K, Harden A, Harris J, Lewin S, Lockwood C (editors), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group. Available from URL: <http://cqrmg.cochrane.org/supplemental-handbook-guidance>

A 'Dual Heritage' for QES

- Hannes, K & Harden, A. (2012) Multi-context versus Context-specific qualitative evidence syntheses: discussing their value for practice and policy decision making processes. *Research Synthesis Methods*, 2(4), 271-278.
- Hannes, K & Lockwood, C (Eds) (2011) *Synthesising Qualitative Research: Choosing the right approach*, Wiley-Blackwell, Australia
- Hannes, K, Lockwood, C, & Pearson, A (2010) A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. *Qualitative Health Res*, 20(12), 1736-43.
- Hannes, K, & Macaitis, K (2012) A move to more systematic and transparent approaches in qualitative evidence synthesis: update on a review of published papers. *Qualitative Research*, 12 (4), **402-442**.
- Harden A (2008) *Critical appraisal and qualitative research: exploring sensitivity analysis*. In: NCRM Research Methods Festival 2008, 30th June - 3rd July 2008, St Catherine's College, Oxford. (Unpublished)
- Harden A, Garcia, J, Oliver, S, Rees, R, Shepherd, J, Brunton, G, & Oakley, A. (2004) Applying systematic review methods to studies of people's views: an example from public health research. *Journal of Epidemiology & Community Health*, 58(9), 794-800.
- Harris, J (2011) Chapter 2: Using qualitative research to develop robust effectiveness questions and protocols for Cochrane systematic reviews. In: Noyes J, Booth A, Hannes K, Harden A, Harris J, Lewin S, Lockwood C (editors), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group. Available from URL <http://cqrmg.cochrane.org/supplemental-handbook-guidance>
- Hind, D, & **Booth, A** (2007) Do health technology assessments comply with QUOROM diagram guidance? An empirical study. *BMC Medical Research Methodology*, 7, 49. Available at: <http://www.biomedcentral.com/1471-2288/7/49>
- Ilott I, **Booth A**, Rick J, & Patterson M (2010) How do nurses, midwives and health visitors contribute to protocol-based care? A synthesis of the UK literature. *International Journal of Nursing Studies*, 47(6):770-780
- Ilott I, Rick J, Patterson M, Turgoose C, & Lacey A. (2006) What is protocol-based care? A concept analysis. *J Nurs Manag*, 14(7):544-52.
- Jones, K. (2004) Mission drift in qualitative research, or moving toward a systematic review of qualitative studies, moving back to a more systematic narrative review.

A 'Dual Heritage' for QES

- The Qualitative Report*, 9(1), 95–112. Retrieved June 12th, 2012, from www.nova.edu/ssss/QR/QR9-1/jones.pdf.
- Jones, L., Atkinson, A., Porcellato, L., Bates, G. et al. (2011) *A systematic review of qualitative research on the views, perspectives and experiences of hepatitis B and C testing among practitioners and people at greatest risk of infection*. Liverpool: Centre for Public Health.
- Jüni, P, Altman, DG, & Egger, M (2001) Assessing the quality of controlled clinical trials. *British Medical Journal*, 323(7303), 42-46.
- Khangura, S, Konnyu, K, Cushman, R, Grimshaw, J, & Moher, D. (2012) Evidence summaries: the evolution of a rapid review approach. *Systematic Reviews* 1(1):10. doi: 10.1186/2046-4053-1-10.
- Kiwanuka SN, Kinengyere AA, Rutebemberwa E, Nalwadda C, Ssengooba F, Olico-Okui, Pariyo GW. (2011) *Dual practice regulatory mechanisms in the health sector: a systematic review of approaches and implementation*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Kvan, T. (2004) The dual heritage of CAAD research, *International Journal of Architectural Computing*, 2 (1), 11-17.
- Laupacis, A & Straus, S. (2007) Systematic reviews: time to address clinical and policy relevance as well as methodological rigor. *Annals of Internal Medicine*, 147(4), 273-4.
- Lee AC, Phillips W, Challen K, & Goodacre S. (2012a) Emergency management in health: key issues and challenges in the UK. *BMC Public Health*, 12(1):884.
- Lee ACK, Challen K, Gardois P, Mackway-Jones K, Carley SD, Phillips W, **Booth A**, Walter D & Goodacre S (2012b) *Emergency planning in health: scoping study of the international literature, local information resources and key stakeholders*. Final report. NIHR Service Delivery & Organisation Programme.
- Lefebvre, C (1994) The Cochrane Collaboration: the role of the UK Cochrane Centre in identifying the evidence. *Health Libraries Review*, 11(4), 235-42.
- Lefebvre, C, Manheimer, E, & Glanville, J (2011) Chapter 6: Searching for studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration. Available from: www.cochrane-handbook.org.
- Lewin, S, Bosch-Capblanch, X, Oliver, S, Akl, EA, Vist, GE, et al (2012) Guidance for evidence-informed decisions about health systems: assessing how much

A 'Dual Heritage' for QES

- confidence to place in health systems evidence. *PLoS Medicine* 9: e1001187.
doi:10.1371/journal.pmed.1001187.
- Lewis, SA, Noyes, J, & Mackereth, S (2010) Knowledge and information needs of young people with epilepsy and their parents: Mixed-method systematic review. *BMC Pediatrics*, 10, 103.
- Liberati, A, Altman, DG, Tetzlaff, J, Mulrow, C, Gøtzsche, PC, Ioannidis, JP, Clarke, M, Devereaux, PJ, Kleijnen, J, & Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*, 339, b2700.
- Lloyd-Jones ML, (2004) Application of systematic review methods to qualitative research: practical issues. *Journal of Advanced Nursing*, 48(3), 271-8.
- Lloyd-Jones ML (2005) Role development and effective practice in specialist and advanced practice roles in acute hospital settings: systematic review and meta-synthesis. *Journal of Advanced Nursing*, 49(2), 191-209.
- Lorenc, T, Pearson, M, Jamal, F, Cooper, C, & Garside, R. (2012) The role of systematic reviews of qualitative evidence in evaluating interventions: a case study. *Research Synthesis Methods* 3, 1–10.
- Major, C & Savin-Baden, M (2010) *An Introduction to Qualitative Research Synthesis: Managing the Information Explosion in Social Science Research*. London: Routledge.
- Major, C. & Savin-Baden, M. (2011) Integration of qualitative evidence: Towards construction of academic knowledge, *Qualitative Research*, 11(6), 1-19.
- Malpass A, Shaw A, Sharp D, Walter F, Feder G, Ridd M, & Kessler D. (2009) "Medication career" or "moral career"? The two sides of managing antidepressants: a meta-ethnography of patients' experience of antidepressants. *Social Science & Medicine*, 68(1):154-68.
- Manning, N (2011) Chapter 8 – Conclusion. In Hannes K and Lockwood C (eds) *Synthesizing Qualitative Research: Choosing the right approach*. London: Wiley Blackwell, 161-172.
- Marshall, IJ, Wolfe, CD, & McKevitt, C (2012) Lay perspectives on hypertension and drug adherence: systematic review of qualitative research. *BMJ*, 345: e3953.
- Martin Hilber A, Kenter E, Redmond S, Merten S, Bagnol B, Low N, & Garside R (2012). Vaginal practices as women's agency in sub-Saharan Africa: a synthesis of meaning and motivation through meta-ethnography. *Social Science & Medicine*, 74(9):1311-23.

A 'Dual Heritage' for QES

- Mays N, Pope C & Popay J (2005) Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research and Policy*, 10(Suppl 1), 6–20.
- Mays N, & Pope C (1996) *Qualitative research in health care*. London: BMJ Publishing Group.
- McLoughlin, D. (1985) A Framework for Integrated Emergency Management. *Public Administration Review*, **45**, 165–172.
- Miles, MB, & Huberman, A.M. (1994) *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd edn. London: Sage Publications.
- Miller, T, Bonas, S, & Dixon Woods, M (2007) Qualitative research on breastfeeding in the UK: a narrative review and methodological reflection. *Evidence & Policy*, 3:197-230
- Moher, D, Weeks, L, Ocampo, M, Seely, D, Sampson, M, Altman, DG, Schulz, KF, Miller, D, Simera, I, Grimshaw, J, & Hoey, J (2011) Describing reporting guidelines for health research: a systematic review. *Journal of Clinical Epidemiology*, 64(7), 718-42
- Moodie C, Hastings G & Thomas J (2011) *Protocol for Plain Tobacco Packaging: A Systematic Review*. Institute of Education Available at: http://phrc.lshtm.ac.uk/papers/PHRC_006_Protocol.pdf Accessed on 28/05/2012
- Murphy, E, Dingwall, R, Greatbatch, D, Parker, S, & Watson, P. (1998) Qualitative research methods in health technology assessment: a review of the literature. *Health Technology Assessment*, 2(16).
- NHS Modernisation Agency and National Institute for Clinical Excellence (2002) *A Step-by-Step Guide to Developing Protocols*. NHS Modernisation Agency and National Institute for Clinical Excellence. Available at: <http://www.qihub.scot.nhs.uk/media/220676/rg0013%20step%20by%20step%20guide%20to%20developing%20protocols.pdf>
- Noblit, GW, & Hare, RD. (1988) *Meta-ethnography: Synthesizing Qualitative Studies*. Newbury Park, CA: Sage Publications.
- Noyes, J & Lewin, S (2011a) Chapter 5: Extracting qualitative evidence. In: Noyes J, Booth A, Hannes K, Harden A, Harris J, Lewin S, Lockwood C (editors), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group. Available from URL <http://cqrmg.cochrane.org/supplemental-handbook-guidance>
- Noyes, J & Lewin, S (2011b) Chapter 6: Supplemental Guidance on Selecting a Method of Qualitative Evidence Synthesis, and Integrating Qualitative Evidence with

A 'Dual Heritage' for QES

- Cochrane Intervention Reviews. In: Noyes J, Booth A, Hannes K, Harden A, Harris J, Lewin S, Lockwood C (editors), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group. Available from URL <http://cqrng.cochrane.org/supplemental-handbook-guidance>
- Noyes, J & Popay J. (2007) Directly observed therapy and tuberculosis: how can a systematic review of qualitative research contribute to improving services? a qualitative metasynthesis. *Journal of Advanced Nursing*, 57: 227-43.
- Noyes, J, Popay, J, Pearson, A, Hannes, K & **Booth, A** on behalf of the Cochrane Qualitative Research Methods group (2008) Qualitative research and Cochrane reviews. In: Higgins, JPT and Green, S (eds) *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: Wiley.
- Oliver, S, Clarke-Jones, L, Rees, R, Milne, R, Buchanan, P, Gabbay J, Gyte G, Oakley A, Stein K (2004) Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach. *Health Technology Assessment*, 8: 1-IV.
- Oliver S, Rees R, Clarke-Jones L, Milne R, Oakley A, Gabbay J, Stein K, Buchanan P, Gyte G (2008) A multidimensional conceptual framework for analysing public involvement in health services research. *Health Expectations*, 11: 72-84
- Oliver, S, Dickson, K, & Newman, M (2012). Getting started with a review. In Gough, D, Oliver, S, & Thomas, J. *An Introduction to Systematic Reviews*. London: Sage Publications.
- Oliver S, Harden A, Rees R, Shepherd J, Brunton G, Garcia J, & Oakley A. (2005) An Emerging Framework for Including Different Types of Evidence in Systematic Reviews for Public Policy. *Evaluation*, 11:428-446.
- Onwuegbuzie AJ & Collins KMT (2007) A typology of mixed methods sampling designs in social science research. *The Qualitative Report* 12(2). Available from: <http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf>.
- Owen J, Carroll C, Cooke J, Formby E, Hayter M, Hirst J, Lloyd-Jones M, Stapleton H, Stevenson M, & Sutton A (2010). School-linked Sexual Health services for Young People (SSHYP): survey and systematic review, *Health Technology Assessment* 2010; 14 (30).
- Papaioannou, D, Cooper, KL, Carroll, C, Hind, D, Squires, H, Tappenden, P, & Logan RF. (2011) Antioxidants in the chemoprevention of colorectal cancer and colorectal

A 'Dual Heritage' for QES

- adenomas in the general population: a systematic review and meta-analysis. *Colorectal Diseases*, 13(10), 1085-99.
- Papaioannou, D, Sutton, A, Carroll, C, **Booth, A**, & Wong, R (2010) Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information & Libraries Journal*, 27(2),114-22.
- Paterson, BL. (2011) "It Looks Great but How do I know if it Fits?": An Introduction to Meta-Synthesis Research, in *Synthesizing Qualitative Research: Choosing the Right Approach* (eds K. Hannes and C. Lockwood), John Wiley & Sons, Chichester.
- Paterson, BL, Thorne, S, Canam, C, & Jillings, C. (2001) *Meta-study of qualitative health research: a practical guide to meta-analysis and metasyntesis*. London: Sage Publications.
- Paterson, BL, Thorne, S, Dewis, M (1998) Adapting to and managing diabetes. *Image Journal of Nursing Scholarship*, 30(1), 57-62.
- Patterson, M, Rick, J, Ilott, I, Lekka, C, Dixon, S & Oluboyede, Y. (2008) *The Contribution of nurses, midwives, and health visitors to protocol-based care and its variants, and the impact of their contribution on patient and staff outcomes, quality and costs of care* (SDO Project 08/1405/079). Report for the National Institute for Health Research Service Delivery and Organisation programme, June 2008.
- Pawson, R (2006a) Digging for Nuggets: How 'Bad' Research Can Yield 'Good' Evidence *International Journal of Social Research Methodology*, 9 (2): 127-142.
- Pawson, R. (2006b) *Evidence-based policy: a realist perspective*. London: Sage Publications.
- Pawson, R. (2006c) Chapter 4 - Realist Synthesis: New Protocols for Systematic Review. In: Pawson R *Evidence-based policy: a realist perspective*. London: Sage Publications, 73-104.
- Pawson, R, Greenhalgh, T, Harvey, G, & Walshe, K (2004) *Realist synthesis: an introduction*. Manchester, UK: ESRC Research Methods Programme.
- Pawson, R, Greenhalgh, T, Harvey, G, & Walshe, K (2005) Realist review—a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, 10 (Suppl 1), 21–34.
- Pearson M, Moxham T, & Ashton K (2011) Effectiveness of search strategies for qualitative research about barriers and facilitators of program delivery. *Evaluation & the Health Professions*, 34(3), 297-308.
- Petticrew, M. (2001). Systematic reviews from astronomy to zoology: myths and misconceptions. *BMJ: British Medical Journal*, 322(7278), 98.

A 'Dual Heritage' for QES

- Petticrew, M, & Roberts, H. (2006) *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell: p. 130.
- Popay J, Rogers A, & Williams G (1998) Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research*, 8, 314–351.
- Pope C, Mays N, & Popay J (2007) Interpretive approaches to evidence synthesis. In Pope, C., Mays, N. & Popay, J, *Synthesising Qualitative and Quantitative Health Evidence: A Guide to Methods*. Maidenhead: McGraw Hill, 72-94.
- Power R (2001) Checklists for improving rigour in qualitative research. Never mind the tail (checklist), check out the dog (research). *BMJ*, 323(7311), 514-515.
- Rambout, L, Hopkins, L, Hutton, B, & Fergusson, D (2007) Prophylactic vaccination against human papillomavirus infection and disease in women: a systematic review of randomized controlled trials. *Canadian Medical Association Journal*, 177(5), 469-479.
- Ring, N, Ritchie, K, Mandava, L, & Jepson, R. (2010) *A guide to synthesising qualitative research for researchers undertaking health technology assessments and systematic reviews*. NHS Quality Improvement Scotland.
- Ritchie, J. & Spencer, L (1994) Qualitative data analysis for applied policy research in A. Bryman and R. G. Burgess [eds.] *Analyzing qualitative data*, [Publisher]: 173-194.
- Rogers, EM. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.
- Rotstein, D, & Laupacis, A (2004) Differences between systematic reviews and health technology assessments: A trade-off between the ideals of scientific rigor and the realities of policy making. *Int J Technol Assess Health Care*, 20, 177-183.
- Sampson, M, McGowan, J, Tetzlaff, J, Cogo, E, & Moher, D. (2008) No consensus exists on search reporting methods for systematic reviews. *Journal of Clinical Epidemiology*, 61(8), 748-754.
- Sampson, M, Tetzlaff, J, & Urquhart, C. (2011) Precision of healthcare systematic review searches in a cross-sectional sample. *Research Synthesis Methods*, 2(2), 119-125.
- Sandelowski, M. (2007) In response to "Brimful of STARLITE". *Journal of the Medical Library Association*, 95(3):233.
- Sandelowski, M, Voils, C, Barroso, J. (2007) Comparability work and the management of difference in research synthesis studies. *Social Science and Medicine*, 64, 236–247.
- Sayers, A (2008) Tips and tricks in performing a systematic review - chapter 4. *British Journal of General Practice*, 58(547), 136.

A 'Dual Heritage' for QES

- Scope A, **Booth A**, & Sutcliffe P (2012) Women's perceptions and experiences of group cognitive behaviour therapy and other group interventions for postnatal depression: a qualitative synthesis. *Journal of Advanced Nursing*, 68(9):1909-19.
- Shaw, RL, **Booth, A**, Sutton, AJ, Miller, T, Smith, JA, Young, B, Jones, DR, & Dixon-Woods, M. (2004) Finding qualitative research: an evaluation of search strategies. *BMC Medical Research Methodology*, 4, 5 Available at: <http://www.biomedcentral.com/1471-2288/4/5>).
- Shepperd, S, Lewin, S, Straus, S, Clarke, M, Eccles, MP, Fitzpatrick, R, Wong, G, Sheikh, A (2009) Can We Systematically Review Studies That Evaluate Complex Interventions? *PLoS Med*, 6(8):e1000086.
- Simera, I, Altman, DG, Moher, D, Schulz, KF, & Hoey, J (2008) Guidelines for reporting health research: the EQUATOR network's survey of guideline authors. *PLoS Med*, 5(6), e139.
- Simera I & Altman DG. (2009) Writing a research article that is "fit for purpose": EQUATOR Network and reporting guidelines. *Evid Based Med*, 14(5), 132-4.
- Squires H, Tappenden P, Cooper K, Carroll C, Logan R, & Hind D. (2011) Cost-effectiveness of aspirin, celecoxib, and calcium chemoprevention for colorectal cancer. *Clinical Therapeutics*, 33(9):1289-305.
- Srivastava, A. & Thomson, SB. (2009) Framework Analysis: a Qualitative Methodology for Applied Policy Research. *JOAAG*, 4 (2). Available at: http://joaag.com/uploads/06_Research_Note_Srivastava_and_Thomson_4_2_.pdf
- Stevenson, M., Scope, A., Sutcliffe, P. A., Booth, A., Slade, P., Parry, G. & Kalthenthaler, E. (2010a). Group cognitive behavioural therapy for postnatal depression: a systematic review of clinical effectiveness, cost effectiveness and value of information analyses. *Health Technology Assessment*, 14(44), 1-135.
- Stevenson MD, Scope A, & Sutcliffe PA (2010b) The cost-effectiveness of group cognitive behavioral therapy compared with routine primary care for women with postnatal depression in the UK. *Value in Health*, 13(5):580-4.
- Subirana, M, Solá, I, Garcia, JM, Gich, I, & Urrútia, G (2005) A nursing qualitative systematic review required MEDLINE and CINAHL for study identification. *Journal of Clinical Epidemiology*, 58(1), 20-5.
- Suri, H (2011) Purposeful Sampling in Qualitative Research Synthesis, *Qualitative Research Journal*, 11(2), 63–75.
- Sweet M, & Moynihan R (2007) *Improving Population Health: The Uses of Systematic Reviews*. Produced by Milbank Memorial Fund, in Collaboration with the Centers

A 'Dual Heritage' for QES

for Disease Control and Prevention (CDC). Available at:

<http://www.milbank.org/uploads/documents/0712populationhealth/0712populationhealth.html>

Thomas, J, & Harden, A. (2007) *Methods for the thematic synthesis of qualitative research in systematic reviews*. NCRM Working Paper. ESRC National Centre for Research Methods. (Unpublished)

Thomas, J, & Harden, A. (2008) Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8(1), 45.

Thomas, J, Harden, A, Oakley, A, Oliver, S, Sutcliffe, K, Rees, R, Brunton, G, & Kavanagh, J. (2004) Integrating qualitative research with trials in systematic reviews. *BMJ*, 328, 1010-1012.

Tong A, Sainsbury P, & Craig J (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*, 19(6):349-57.

Tong A, Flemming K, McInnes E, Oliver S, & Craig J (2012a) *Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ*. Presentation at Cochrane Colloquium, Auckland, New Zealand, October 2012.

Tong A, Flemming K, McInnes E, Oliver S, & Craig J (2012b) Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology* 12:181. doi: 10.1186/1471-2288-12-181.

Tranfield, D, Denyer, D, & Smart, P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Brit J Management*, 14(3), 207-22.

Urquhart, C. (2010) Systematic reviewing, meta-analysis and meta-synthesis for evidence-based library and information science. *Information Research*, 15(3) colis708.

[Available at <http://InformationR.net/ir/15-3/colis7/colis708.html>]

Watt A, Cameron A, Sturm L, Lathlean T, Babidge W, Blamey S, et al (2008). Rapid versus full systematic reviews: validity in clinical practice? *Australian & New Zealand Journal of Surgery*, 78:1037-1040.

Weed, M (2005) "Meta interpretation": a method for the interpretive synthesis of qualitative research. *Forum: Qualitative Social Research*, 6(1), art.37. Available from: <http://www.qualitative-research.net/fqs-texte/1-05/05-1-37-e.htm>.

Weir MC, Grimshaw JM, Mayhew A, & Fergusson D (2012) Decisions about lumping vs. splitting of the scope of systematic reviews of complex interventions are not well

A 'Dual Heritage' for QES

- justified: a case study in systematic reviews of health care professional reminders. *Journal of Clinical Epidemiology*, 65(7):756-63.
- Wells, K & Littell, JH (2009) Study quality assessment in systematic reviews of research on intervention effects. *Research on Social Work Practice*, 19, 52-62.
- Wilkinson, A, Papaioannou, D, Keen, C, & **Booth, A** (2009) The role of the information specialist in supporting knowledge transfer: a public health information case study. *Health Information & Libraries Journal*, 26(2), 118-25.
- Williams, B (2001) Checklists for improving rigour in qualitative research. Including personal reflections might help. *BMJ*, 323(7311), 515.
- Windle G, Bennett KM, & Noyes J (2011) A methodological review of resilience measurement scales. *Health Quality of Life Outcomes*. 9:8.
- Wolcott, HF (2001) *Writing up Qualitative Research*. Thousand Oaks: Sage Publications.
- Zhang, L, Sampson, M, & McGowan, J (2006) Reporting of the Role of the Expert Searcher in Cochrane Reviews." *Evidence Based Library and Information Practice* [Online], 1 (4), 3-16. Web. 28 May. 2012
- Zhao, S. (1991) Metatheory, metamethod, meta-data-analysis: What, why, and how? *Sociological Perspectives*, 34, 377-390.

A 'Dual Heritage' for QES

Appendix A – Summary of Permissions

All permissions for reproduction of articles have been verified using the official JISC (Joint Information Systems Committee)-funded ROME0 <http://www.sherpa.ac.uk/romeo/> resource on **Publisher copyright policies & self-archiving**, used by the *White Rose Repository* and other academic repositories to provide authoritative guidance on permitted versions.

Reference	Permitted Versions	Version Included	Author Permissions Received
Paper M1. Booth A. "Brimful of STARLITE": toward standards for reporting literature searches. <i>J Med Libr Assoc.</i> 2006 Oct; 94(4): 421-429	Author can archive publisher's version/PDF	Publisher's Version pdf	Not Applicable
Paper M2. Dixon-Woods, M, Booth, A & Sutton AJ. Synthesizing qualitative research: a review of published reports <i>Qualitative Research</i> 2007 Aug 7: 375-422.	Post-print version with changes from referees comments can be used, Publisher copyright and source must be acknowledged	Post print version	Dixon-Woods Sutton
Paper M3. Carroll C, Booth A, Lloyd-Jones M. Should we Exclude Inadequately-reported Studies from Qualitative Systematic Reviews? An Evaluation of Sensitivity Analyses in Two Case Study Reviews, <i>Qualitative Health Research</i> 2012; 22 (10): 1425-1434.	Subject to 12 month embargo author can archive post-print (i.e final draft post-refereeing)	Post print version. Embargo until October 2013.	Carroll Lloyd-Jones
Paper M4. Carroll C, Booth A, Cooper K. A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents. <i>BMC Med Res Methodol.</i> 2011 Mar 16;11:29.	Author can archive publisher's version/PDF	Publisher's Version pdf	Carroll Cooper
Paper M5. Booth A, Carroll C, Iltott I, Low LL & Cooper K. Desperately Seeking Dissonance: Identifying the "Disconfirming Case" in Qualitative Evidence Synthesis. <i>Qualitative Health Research</i> 2013; 23 (1): 126-141.	Subject to 12 month embargo author can archive post-print (i.e final draft post-refereeing)	Post print version. Embargo until January 2014.	Carroll Iltott Low Cooper

Appendix B - Paper M1

This is the publishers' version. Published source is acknowledged as:

**Booth A. "Brimful of STARLITE": toward standards for reporting literature searches.
J Med Libr Assoc. 2006 Oct; 94(4):421-429.**

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1629442/>

"Brimful of STARLITE": toward standards for reporting literature searches

Andrew Booth, MSc, MCLIP, Reader in Evidence Based Information Practice and Director of Information Resources,
A.Booth@sheffield.ac.uk
School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street,
Sheffield, S1 4DA, United Kingdom

Context: Systematic reviews of qualitative research studies extend understanding of health care beyond effectiveness to acceptability and user views.

Objective: The paper surveys reports of qualitative systematic reviews and, by characterizing techniques used to identify articles for inclusion, proposes standards for reporting of literature searches.

Data Sources and Study Selection: A search of MEDLINE was performed for qualitative systematic reviews published from 1988 to December 2004, supported by searches of CINAHL, Web of Knowledge (including the Science and Social Sciences Citation Index), and the Cochrane Methodology Register, and Internet searches using the Copernic Agent Professional meta-search agent. Studies were included if they used techniques of qualitative synthesis in reviewing research studies in health care. Narrative reviews were excluded.

Data Extraction: Authors, year of publication, sampling strategy, databases, keywords, and other approaches used were extracted.

Data Synthesis: Sixty-four studies were identified, and forty-three met inclusion criteria for this review. A summary of searching methods was produced and used to construct the STARLITE mnemonic (sampling strategy, type of study, approaches, range of years, limits, inclusion and exclusions, terms used, electronic sources).

Conclusions: Considerable variation exists in search methods for qualitative systematic reviews. While diversity in methods is appropriate during the development of review methodology, major concerns remain about the absence of an accepted standard and the consequent poor quality of reporting.

BACKGROUND

Recent years have seen increasing recognition of the potential contribution of qualitative health research to informing health policy and clinical practice [1, 2]. Qualitative approaches allow the researcher to explore the richness and complexity of human experience in a given context [3]. In representing human experience, this type of research provides important information about such aspects as the appropriateness of care and the impact of illness. As such, it complements the role of quantitative research where the focus is often on improved understanding of the effectiveness of health care. Qualitative research may also give consumers "a voice in the decision-making process through the documentation of their experiences, preferences, and priorities" [2]. Specifically, in health technology assessment (HTA), a properly employed qualitative approach can provide "valuable information on the implementation and impact of health technologies on both health professionals and patients" [4]. In particular, the value of systematic reviews of qualitative research that synthesize the findings of multiple studies

Highlights

- Systematic reviews of qualitative research studies are limited by poor quality reporting of search methods.
- Standards for reporting literature searches must acknowledge the demands of both quantitative *and* qualitative systematic reviews.
- The mnemonic STARLITE (sampling strategy, type of study, approaches, range of years, limits, inclusion and exclusions, terms used, electronic sources) may be used to convey the essential elements for reporting literature searches.

Implications

- There is a pressing need to achieve international consensus about standards for reporting literature searches.
- Further work needs to define the contents of each proposed element and the ways these elements are to be evaluated.
- Librarians have a key role in defining standards for systematic reviews and their subsequent reporting.



Supplemental electronic content is included with this paper on PubMed Central.

covering the same topic is being increasingly acknowledged.

Systematic reviews—in which evidence has been systematically identified from comprehensive searches of the published and unpublished literature, appraised for quality, and then synthesized to produce generalizable messages—have become a key tool in the development of an evidence base [5]. A qualitative systematic review, also referred to as a qualitative meta-synthesis, is a method for integrating or comparing the findings from qualitative studies [6]. The accumulated knowledge resulting from this process may lead to the development of a new theory, an overarching “narrative,” a wider generalization, or an “interpretative translation” [7]. Whereas a quantitative meta-synthesis, or meta-analysis, aims to pool the numerical results of individual quantitative studies, a qualitative meta-synthesis looks for “themes” or “constructs” that lie in or across individual qualitative studies. The goal of such a qualitative meta-synthesis is not aggregative in the sense of “adding studies together,” as with a meta-analysis. On the contrary, it is interpretative in broadening understanding of a particular phenomenon [8]. Within this broader category of “qualitative meta-synthesis,” the narrow term “meta-ethnography” [9] refers to the specific method of data synthesis that has been most widely adopted in the literature to date. For example, Paterson and colleagues have identified thirty-eight studies examining the firsthand experience of living with diabetes [10]. They have found that the prevailing metaphor was the concept of balance and specific subthemes identified across multiple studies included “knowing one’s body,” “learning how to manage diabetes,” and “fostering supportive, collaborative relationships with others.”

Compared to systematic reviews of the quantitative research literature, qualitative systematic reviews in health care are a much more recent phenomenon. Whereas quantitative reviews are conducted according to such guidelines as the Cochrane Handbook [11] and the National Health Service (NHS) Centre for Reviews and Dissemination report number 4 [12], similar accepted principles for qualitative reviews are lacking [13]. The second edition of published NHS Centre for Reviews and Dissemination guidelines does cater superficially to qualitative research in the context of effectiveness research, but it is widely acknowledged that qualitative reviews may be used for a much broader range of purposes, which may or may not include the “typical” effectiveness question [12]. Included in a quantitative legacy that is being increasingly challenged by qualitative systematic reviewers are such acknowledged systematic review mechanisms as checklists [6, 14–16] and a hierarchy of evidence [17].

Notwithstanding considerable progress, vigorous debate exists as to whether it is appropriate to apply conventional systematic review techniques, developed primarily for quantitative systematic reviews and meta-analyses, to reviews of qualitative research [18]. Some commentators argue that it is more appropriate to develop and apply methods analogous to those

used in conducting primary qualitative research, employing familiar techniques such as purposive sampling and theoretical saturation in preference to quantitative-centric review methods [19, 20]. Regardless of one’s stance on such issues, applying such methods to qualitative research presents significant philosophical and practical challenges [21].

This comparative immaturity of methods for qualitative systematic reviews is mirrored in the specific context of the identification of studies. Whereas techniques for retrieval of quantitative study designs (such as randomized controlled trials) are relatively far advanced, it is only comparatively recently that attention has started to be focused on methods for identifying qualitative research studies [2, 22–24]. This deficiency is fittingly signaled by the fact that a chapter on searching for qualitative research has not yet been included in the international HTA community’s otherwise impressively comprehensive *E-text on Health Technology Assessment Information Resources* [25].

This investigation has been conducted to accompany and inform methodological advances pursued via a UK Economic and Social Research Council (ESRC)–funded† project on approaches to synthesizing qualitative and quantitative research. This study aims to identify published examples of qualitative systematic reviews in health care published between 1988 and December 2004. It then seeks to characterize these reviews with regard to methods used for the sampling of studies for inclusion in each review and the search techniques used to identify such studies. In doing so, this study attempts to outline priorities for future development of the methods of qualitative systematic reviews, particularly with respect to standards for the identification and reporting of included studies.

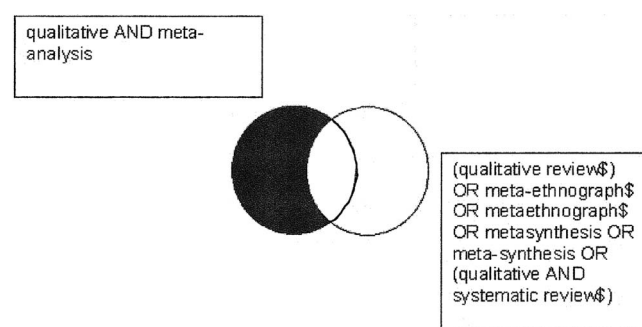
The nature of the study question together with the resource limitations of an unfunded project mean that it is not possible to conduct a full systematic review of identified qualitative systematic reviews. The methodology used for this study is a systematic survey of the literature. It employs systematic searching methods to identify qualitative systematic reviews. However, it makes no attempt to deliver judgments on the quality of retrieved studies or to validate independently decisions about the inclusion or exclusion of retrieved studies. It may thus be considered an “epidemiological survey” in attempting to quantify the absence or presence of key characteristics of the literature.

METHODS

As indicated in the description of a qualitative systematic review above, a major issue for this systematic survey of the literature is the variation that exists with regard to the terminology used to describe qualitative systematic reviews. This variation applies to “false-negatives”; that is, reviews that use qualitative methods and yet are not identifiable by such terms as “qualitative systematic review,” “meta-ethnography,”

† Principal investigator: Mary Dixon-Woods, project ID H333250043.

Figure 1
Verification strategy



or "meta-synthesis." For example, this author was involved in one such study described simply as a "systematic review" [26]. However, it also applies to "false positives"; that is, to reviews that claim to be "qualitative systematic reviews" but simply use the term to differentiate from meta-analyses, that is, quantitative systematic reviews. For example, there are frequent instances of qualitative systematic reviews of randomized controlled trials. Such a situation is further compounded because few journals in which qualitative systematic reviews are published utilize structured abstracts. Such abstracts can help to clarify whether systematic searches have taken place and whether an established method of synthesis (such as Noblit and Hare's meta-ethnography [9]) has been employed.

A comprehensive search was undertaken of the social science, health, and information science literature using PubMed, MEDLINE, CINAHL, Web of Knowledge (including the Science and Social Sciences Citation Index), and the Cochrane Methodology Register. A sensitive search strategy was used combining the keywords "(qualitative review\$) OR meta-ethnograph\$ OR metaethnograph\$ OR metasynthesis OR meta-synthesis OR (qualitative AND systematic review\$)." Although the combination of "qualitative AND meta-analysis" appeared to yield some relevant hits, it greatly increased the retrieval of manifestly irrelevant records. A verification strategy, produced by searching for this combination and then using the Boolean "NOT" with the sensitive search strategy, revealed no unique relevant records (Figure 1). As it appeared that indexers were adding the index term "meta-analysis" to recognize the presence in the title or abstract of the term "meta-synthesis," the final strategy as given was considered adequate.

To be included, a study had to meet two criteria: the study, or at the very least a significant part of it, should have been a qualitative systematic review that reports search strategies and/or techniques, and it should have been conducted in a health care context. All decisions on inclusion were made by the author. In cases of doubt, reference was made to the full-text of the article in question. A list of potentially eligible and yet excluded studies was maintained to assist transparency. Citations to key qualitative methodological texts

were followed up, and Related Articles features on MEDLINE and Web of Knowledge were also utilized. Searches of the Internet using the Copernic Agent Professional meta-search agent <<http://www.copernic.com>> were used to track down unpublished and gray literature. This software is particularly useful for systematic literature reviews, because it searches multiple search engines (including AlltheWeb, Alta Vista, Hotbot, Lycos, and Yahoo), saves the results in an "audit trail" on the searcher's computer, and allows manual weeding of results for inclusion or exclusion. Searches covered the period from 1988, when a key text on metasynthesis was published [9], to December 2004 and were restricted to English language, due to practical constraints. There were no other restrictions.

In excess of 400 articles passed an initial screening by title. Review of abstracts and text from those articles yielded 64 publications that met inclusion criteria. Published examples of meta-syntheses were obtained, and data regarding search methods were extracted into a matrix. A summary produced from this matrix is seen in Table 1. Key variables included authors, year of publication, the sampling strategy (including inclusion and exclusion criteria), and the search strategy (including databases, search terms used, and other searching approaches). The sampling strategy was defined as "comprehensive" if the search strategy attempted to retrieve all relevant studies in a topic area in a conventional systematic review manner, "purposeful" if the search strategy followed an underlying rationale in only trying to find a subset of predefined studies, and "opportunistic" if a review worked with a convenience subset of already retrieved studies. For limits, where languages for inclusion were reported, studies were categorized according to whether the reviews covered "English only," "other (specified) languages," or "all languages" (i.e., no language restrictions). For ease of presentation, all other categories of data extraction are simply coded in Table 1 as "Yes" for item reported or "No" for item not reported. However, full details of strategies were recorded in a matrix constructed as a tool for data extraction. Data on the topic area were also recorded, although only methodological data were reported in this survey.

RESULTS

Overall, sixty-five studies were identified; of these, forty-four (68%) reported at least one of three elements (databases, keywords, other approaches) of their search methods (Table 1) [10, 21, 26–67]. The remaining twenty-one studies did not report any elements of their search strategies (Table 2; find online) [7, 28, 68–88]. Four of these referred to another publication for details of their methods. Of the remaining seventeen, twelve studies did not attempt to conduct a systematic search of the literature, four simply represented an analysis of papers previously produced by the authors, and the remaining one analyzed studies from a previous report. A median of five databases was used in each review (range 1 to 23). The largest number of

Table 1

Included studies with level of literature search reporting (n = 44)

Authors (Year)	Sampling strategy	Study type	Approaches	Range of years	Limits	Inclusions and exclusions	Terms used	Electronic sources
Attree (2004) [27]	Comprehensive	No	Yes	Yes	English	No	No	No
Barroso et al. (2003) [28]	Purposive	No	Yes	Yes	English	Yes	Yes	Yes
Beck (2001) [29]	Comprehensive	No	No	Yes	English	No	Yes	Yes
Beck (2002) [30]	Comprehensive	Yes	Yes	Yes	English	No	Yes	Yes
Beck (2002) [31]	Comprehensive	No	Yes	Yes	English	No	No	Yes
Beverley et al. (2004) [26]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
Brauer et al. (2001) [32]	Comprehensive	No	No	Yes	English	No	Yes	Yes
Burke et al. (1998) [33]	Comprehensive	No	No	Yes	English	No	Yes	Yes
Campbell et al. (2003) [21]	Purposive	No	Yes	Yes	English	Yes	Yes	Yes
Carroll (2004) [34]	Comprehensive	No	No	Yes	English	No	Yes	Yes
Chapple and Rogers (1999) [35]	Comprehensive	Yes	Yes	Yes	English	No	Yes	Yes
Clemmens (2003) [36]	Purposive	No	No	Yes	English	Yes	Yes	Yes
Cook et al. (2001) [37]	Opportunistic	No	Yes	Yes	English	No	No	Yes
Duggan and Banwell (2004) [38]	Comprehensive	No	Yes	Yes	English	No	No	Yes
Evans and FitzGerald (2002) [39]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
Fingfeld (1999) [40]	Comprehensive	No	No	Yes	English	No	No	Yes
Fingfeld (2000) [41]	Comprehensive	No	No	Yes	English	No	Yes	Yes
Fredriksson (1999) [42]	Purposive	No	Yes	Yes	English/Scand*	No	Yes	Yes
Frederiksson (2001) [43]	Purposive	No	Yes	Yes	English/Scand	No	Yes	Yes
Garcia et al. (2002) [44]	Comprehensive	No	Yes	No	English	No	No	Yes
Jones (2004) [45]	Comprehensive	No	Yes	Yes	All	No	Yes	Yes
Kearney (2001) [46]	Purposive	Yes	Yes	Yes	English	No	No	Yes
Lefler and Bondy (2004) [47]	Comprehensive	No	Yes	Yes	All	Yes	Yes	Yes
Lemmer et al. (1999) [48]	Comprehensive	No	No	No	English	No	Yes	Yes
McEwan et al. (2004) [49]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
McKevitt et al. (2004) [50]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
McNaughton (2000) [51]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
Mold et al. (2003) [52]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
Murray et al. (2003) [53]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
Neill (2000) [54]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
Nelson (2002) [55]	Comprehensive	No	No	Yes	English	No	No	Yes
Nelson (2003) [56]	Comprehensive	Yes	No	No	English	Yes	Yes	Yes
Paterson (2001) [57]	Comprehensive	No	Yes	Yes	English	No	No	Yes
Paterson et al. (1998) [10]	Comprehensive	No	Yes	Yes	English	No	No	Yes
Roberts et al. (2002) [58]	Comprehensive	Yes	Yes	Yes	English	No	No	Yes
Rogers (1997) [59]	Comprehensive	Yes	Yes	Yes	English	Yes	Yes	Yes
Rowe and Rudkin (1999) [60]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes
Sandelowski and Barroso (2003) [61]	Comprehensive	Yes	No	Yes	English	Yes	No	No
Sandelowski and Barroso (2003) [62]	Comprehensive	Yes	No	Yes	English	Yes	No	No
Sydes et al. (2004) [63]	Comprehensive	No	Yes	Yes	All	No	No	Yes
Thorne and Paterson (1998) [64]	Comprehensive	No	Yes	Yes	English	No	No	Yes
Thorne et al. (2002) [65]	Comprehensive	No	Yes	Yes	English	No	No	Yes
Walter et al. (2004) [66]	Comprehensive	Yes	Yes	No	All	Yes	Yes	Yes
Woodward and Webb (2001) [67]	Comprehensive	No	Yes	Yes	English	No	Yes	Yes

* Scand = Scandinavian languages.

Table 3

Elements of the STARLITE mnemonic

Element	Explanatory notes
S: Sampling strategy	<ul style="list-style-type: none"> ■ Comprehensive: attempts to identify all relevant studies on the topic ■ Selective: attempts to identify all relevant studies but only within specified limits ■ Purposive: samples from specific disciplines, years, journals
T: Type of studies	<ul style="list-style-type: none"> ■ Fully reported: describes actual study types (e.g., grounded theory) or designs to be included ■ Partially reported: uses an "umbrella" category such as "qualitative studies" without defining what this means
A: Approaches	<ul style="list-style-type: none"> ■ Approaches other than electronic subject searches (see below) ■ Example: hand-searching ■ Citation snowballing
R: Range of years (start date–end date)	<ul style="list-style-type: none"> ■ Fully reported: includes start and end dates with justification for time period chosen ■ Partially reported: includes start and end dates but only determined available coverage of databases
L: Limits	<ul style="list-style-type: none"> ■ Functional limits that are applied for logistic reasons but do not alter the topic conceptually (e.g., human, English etc.)
I: Inclusion and exclusions	<ul style="list-style-type: none"> ■ Conceptual limitations that mediate the scope of the topic area (e.g., geographical location, setting, or a specific focus of study)
T: Terms used	<ul style="list-style-type: none"> ■ Fully present: example of a sample search strategy from one or more of the main databases ■ Partially present: reports terminology used but without evidence of search syntax and operators
E: Electronic sources	<ul style="list-style-type: none"> ■ Reports databases used and, optimally, search platforms and vendors to assist in replication

databases was searched in studies that exclusively involved information professionals in the review process [26, 38].

Keywords (search terms used) were only reported in twenty-five studies. While no formal analysis of the quality of search strategies was undertaken, certain characteristics could be observed from the literature. There was little evidence for the use of techniques to maximize retrieval such as truncation or explosion of subject terms. Where search terms were reported, it was not usually clear whether they were free-text terms or approved subject headings. Studies did not typically present a sample search strategy, not even as an appendix, and thus it was difficult to observe whether Boolean operators had been employed correctly. In addition, it was unclear whether methodological terms had been used in addition to subject-specific terms to privilege qualitative studies (e.g., the term "qualitative"). Only nine of the studies reported keywords specifically to retrieve qualitative studies. No formal data were extracted on whether a librarian or information specialist was involved in the search process. However, very few studies acknowledged such a contribution in either the authorship or the acknowledgements, and, furthermore, most studies failed to report any such involvement in their methods sections.

The database most frequently mentioned was CINAHL (31 times). MEDLINE was mentioned thirty times, the PsycINFO/PsycLIT/Psychological Abstracts combination was mentioned twenty-five times, and Sociological Abstracts/SOCIOFILE (9 times) and ERIC (5 times) were also listed.

The most common supplementary strategies used alongside searching bibliographic databases were following up reference lists (17 times) and hand-searching (13 times). The most comprehensive report of the literature searching process itemized 23 databases, 10 Websites, and 6 other techniques and reported keywords [26].

Although this study focused on the methodological content of included studies, not their topical content, it was interesting to observe that the most common topics related to chronic disease and to women's health. This observation, which requires further empirical exploration, attests to the possibility that qualitative systematic reviews are being utilized as a facilitative method to provide a "collective voice" to disenfranchised groups [89].

In performing the data extraction, a necessary prelude to reporting the presence or absence of certain characteristics of literature-searching approaches in qualitative systematic reviews, the author has identified several features that might usefully form the basis for future reporting standards for literature searches for systematic reviews. These are briefly outlined below.

PROPOSED STANDARDS FOR REPORTING LITERATURE SEARCHES

The systematic review movement has already driven the development of standards to improve the quality

of reporting of quantitative systematic reviews (e.g., QUOROM [90] and MOOSE [91]). Such standards make it easier for readers to assess the quality of such reviews and for researchers to replicate their methods [92]. To date, no standards have been published for reporting of literature searches, so critical to the successful conduct of systematic reviews.

Clearly, standards need to cover all stages of the systematic review process. Nevertheless, the case is particularly strong for focusing on the quality of reporting methods for identification of included studies, at least in the first instance. Reasons for this include the poor quality of reporting as observed by this survey, the impact of sampling decisions on the findings of qualitative reviews, and the fact that, once a question has been identified, decisions made at this stage determine the remainder of the review process. Table 3 encapsulates a proposed framework for reporting the quality of literature searches based on the empirical findings from this review and supported by the author's extensive experience conducting other forms of synthesis such as HTAs, guidelines, and quantitative systematic reviews. The elements to be included when reporting literature searching to allow a reader to assess the quality of a search and to replicate it, if necessary, are conveyed using the mnemonic STARLITE.

STARLITE constitutes not simply a memorable mnemonic but also serves as an acronym for **Standards for Reporting Literature** searches. Many of the above elements are already widely acknowledged as important in the recording of literature searches, although no formal standards exist. A notable addition, however, is the inclusion of "sampling strategy." This addition is stimulated by the specific needs of qualitative systematic reviews, where an assumption of comprehensiveness cannot be made. It is possible for a qualitative systematic review to be explicit and systematic, while not aspiring to comprehensiveness, if it employs purposive sampling of the literature from certain disciplines or even from particular years. Indeed closer examination of many quantitative systematic reviews reveals that they create an illusion of "comprehensiveness," when the reality is that the studies that they actually include are shaped by arbitrary decisions about search strategies dictated themselves by time and resource limitations.

The author intends, subject to agreement with colleagues in the international HTA community, this approach to form a framework for accepted standards for reporting of literature searches for HTAs and systematic reviews in general. An outline example of reporting for a qualitative systematic review using the STARLITE framework is provided (Table 4; find online). Within this framework, future work could concentrate on specifying how exactly the contents of each element should be reported and subsequently evaluated. The magnitude of the task ahead is emphasized by returning to the survey dataset and summarizing the presence or absence of the STARLITE elements (Table 5).

Table 5

Completeness of reporting (n = 44)

	Fully or partially present (percentage)	Absent (percentage)
Sampling strategy	44 (100)	0 (—)
Type of study	9 (20)	35 (80)
Approaches	32 (73)	12 (27)
Range of years	40 (91)	4 (9)
Limits (e.g., English)	44 (100)	0 (—)
Inclusion and exclusions	9 (20)	35 (80)
Terms used	28 (64)	16 (36)
Electronic sources	41 (93)	3 (7)

DISCUSSION

This literature survey has sought to identify the presence and absence of prespecified characteristics in a body of literature believed to meet an operational definition of a "qualitative systematic review" in health care. It is recognized that use of a single reviewer for judgments on inclusion and exclusion of studies opens the possibility of bias. This limitation would indeed be serious if the reviewer had subsequently attempted to make value judgments on whether or not the literature searching methods are "adequate" or not. Such an approach would require the addition of several quality procedures such as the use of additional reviewers, a quality checklist, and assessment of inter-rater reliability. By retaining a descriptive, rather than evaluative, focus the author has been able to characterize a population of qualitative systematic review studies. This population does, of course, have the further potential to be extended further through more exhaustive search procedures such as hand-searching of key qualitative journals such as *Qualitative Research and Qualitative Health Research*.

Further work is required on assessing the quality of the reported search methods and evaluating their likely impact on the quality of the subsequent reviews. In this context, it is interesting to observe commencement of the *Evaluating Health Technology Assessment Searches* (EHTAS) research study to develop a quality assessment checklist for searches used in HTAs and systematic reviews. Finally, this survey has been independently examined in a wider study investigating the quality of reporting of the methods of all stages of the qualitative systematic review, not just the searching, and there are plans to use this data set in further methodological research.

The field of qualitative systematic review is still relatively immature, with no consensus yet on what constitutes such a review. This is illustrated by examples identified from the tables above; does searching only one database, synthesizing a small number of studies by a select group of authors, or identifying studies systematically but stopping short of synthesis rightly constitute a qualitative systematic review? Of course, each criterion exists on a continuum, rendering judgments on what should be included or excluded in this genre necessarily subjective. Ironically, this means that this survey, while possessing a measure of systematicity, falls short of the requirement to be easily reproducible.

CONCLUSION

Recent years have seen ongoing improvement in methods of conducting and reporting systematic reviews. While these two issues remain discrete stages of the review process, they are interconnected; one cannot automatically conclude that a poorly reported review has been badly conducted, but such poor reporting does mean that many of the intrinsic virtues of a systematic review are negated. For example, systematic reviews draw strength from the fact that they claim to be both explicit and reproducible. In 1987, Mulrow [93] highlighted the poor state of the medical review article. Over a decade later, in 1999, McAlister and colleagues [94] found that less than a quarter of published reviews described how evidence was identified, evaluated, or integrated.

Librarians have a key role in the further development of systematic review methods, particularly as they relate to retrieval of the evidence. This is true for both the conduct of the review itself and its subsequent reporting. Findings from this survey have the potential to inform the future work of groups of international information specialists, such as the Cochrane Collaboration Information Retrieval Methods Group [95], in developing standards for reporting literature searches to the benefit of both quantitative and qualitative systematic reviews. In the meantime, these findings are offered with the exhortation that authors and editors improve the quality of methods for identifying studies for inclusion in systematic reviews. It is hoped that this article, and its underpinning research, will stimulate improvements in conducting and reporting systematic reviews, both qualitative and quantitative.

ACKNOWLEDGMENTS

The author thanks his collaborators on Economic and Social Research Council Research Methods Programme project number: H333250043 for providing the stimulus for this spin-off project, especially Mary Dixon-Woods who put the team together.

REFERENCES

1. RITCHIE JE. Using qualitative research to enhance the evidence-based practice of health care providers. *Aust J Physiother* 1999;45(4):251-6.
2. EVANS D. Database searches for qualitative research. *J Med Libr Assoc* 2002 Jul;90(3):290-3.
3. MADJAR I, TAYLOR B, LAWLER J. The role of qualitative research in evidence based practice. *Collegian* 2002 Oct;9(4):7-9.
4. MURPHY E, DINGWALL R, GREATBATCH D, PARKER S, WATSON P. Qualitative research methods in health technology assessment: a review of the literature. *Health Technol Assess* 1998;2(16):iii-ix,1-274.
5. TRANFIELD D, DENYER D, SMART P. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Brit J Management* 2003;14(3):207-22.
6. SANDELOWSKI M, DOCHERTY S, EMDEN C. Focus on qual-

- itative methods. qualitative metasynthesis: issues and techniques. *Res Nurs Health* 1997 Aug;20(4):365-71.
7. BECK CT. Seeing the forest for the trees: a qualitative synthesis exercise. *J Nurs Educ* 2003 Jul;42(7):318-23.
 8. JENSEN LA, ALLEN MN. Meta-synthesis of qualitative findings. *Qual Health Res* 1999;6(4):553-60.
 9. NOBLIT GW, HARE RD. Meta-ethnography: synthesizing qualitative studies. Newbury Park, CA: Sage, 1988.
 10. PATERSON BL, THORNE S, DEWIS M. Adapting to and managing diabetes. *Image J Nurs Sch* 1998;30(1):57-62.
 11. HIGGINS JPT, GREEN S, eds. Cochrane handbook for systematic reviews of interventions 4.2.5. [Web document]. Cochrane Collaboration. [cited 15 Sep 2005]. <<http://www.cochrane.org/resources/handbook/hbook.htm>>.
 12. NATIONAL HEALTH SERVICE CENTRE FOR REVIEWS AND DISSEMINATION. Undertaking systematic reviews of research on effectiveness. CRD's guidance for those carrying out or commissioning reviews. CRD report number 4. 2nd ed. [Web document]. York, UK: University of York, 2001. [cited 4 Jan 2006]. <<http://www.york.ac.uk/inst/crd/report4.htm>>.
 13. DIXON-WOODS M, FITZPATRICK R. Qualitative research in systematic reviews has established a place for itself. *BMJ* 2001 Oct 6;323(7316):765-6.
 14. BARBOUR RS. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *BMJ* 2001 May 5;322(7294):1115-7.
 15. EAKIN JM, MYKHALOVSKIY E. Reframing the evaluation of qualitative health research: reflections on a review of appraisal guidelines in the health sciences *J Eval Clin Pract*. 2003 May;9(2):187-94.
 16. DIXON-WOODS M, SHAW RL, AGARWAL S, SMITH JA. The problem of appraising qualitative research. *Qual Saf Health Care* 2004 Jun;13(3):223-5.
 17. CESARIO S, MORIN K, SANTA-DONATO A. Evaluating the level of evidence of qualitative research. *J Obstet Gynecol Neonatal Nurs* 2002 Nov-Dec;31(6):708-14.
 18. BOOTH A. Cochrane or cock-eyed? how should we conduct systematic reviews of qualitative research? [Web document]. Presented at: Qualitative Evidence Based Practice Conference; University of Coventry; 14-16 May 2001. [cited 4 Jan 2006]. <<http://www.leeds.ac.uk/educol/documents/00001724.htm>>.
 19. WEED M. "Meta interpretation": a method for the interpretive synthesis of qualitative research. *Forum: Qualitative Soc Res* [serial online]. 2005 Jan;6(1),art. 37. [cited 4 Jan 2006]. <<http://www.qualitative-research.net/fqs-texte/1-05/05-1-37-e.htm>>.
 20. PAWSON R, GREENHALGH T, HARVEY G, WALSHE K. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005 Jul;(10 suppl 1):21-34.
 21. CAMPBELL R, POUND P, POPE C, BRITTEN N, PILL R, MORGAN M, DONOVAN J. Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Soc Sci Med*. 2003 Feb;56(4):671-84.
 22. SHAW RL, BOOTH A, SUTTON AJ, MILLER T, SMITH JA, YOUNG B, JONES DR, DIXON-WOODS M. Finding qualitative research: an evaluation of search strategies. *BMC Med Res Methodol* 2004 Mar 16;4:5.
 23. SHAW RL, BOOTH A, SUTTON AJ, MILLER T, SMITH JA, YOUNG B, JONES DR, DIXON-WOODS M. Electronic literature search for systematic review of qualitative evidence. *J Epid & Community Health* 2003;57(suppl 1):A15.
 24. GRANT MJ. How does your searching grow? a survey of search preferences and the use of optimal search strategies in the identification of qualitative research. *Health Info Libr J* 2004 Mar;21(1):21-32.
 25. TOPFER LA, AUSTON I, eds. E-text on health technology assessment (HTA) information resources. [Web document]. Bethesda, MD: National Library of Medicine. [updated Nov 2005; cited 4 Jan 2006]. <<http://www.nlm.nih.gov/nichsr/ehta/>>.
 26. BEVERLEY CA, BATH PA, BOOTH A. Health information needs of visually impaired people: a systematic review of the literature. *Health Soc Care Community* 2004 Jan;12(1):1-24.
 27. ATTREE P. Growing up in disadvantage: a systematic review of the qualitative evidence. *Child Care Health Dev* 2004 Nov;30(6):679-89.
 28. BARROSO J, GOLLOP CJ, SANDELOWSKI M, MEYNELL J, PEARCE PF, COLLINS LJ. The challenges of searching for and retrieving qualitative studies. *West J Nurs Res* 2003 Mar;25(2):153-78.
 29. BECK CT. Caring within nursing education: a metasynthesis. *Nurs Educ* 2001 Mar;40(3):101-9.
 30. BECK CT. Postpartum depression: a metasynthesis. *Qual Health Res* 2002 Apr;12(4):453-72.
 31. BECK CT. Mothering multiples: a meta-synthesis of qualitative research. *MCN Am J Matern Child Nurs* 2002 Jul-Aug;27(4):214-21.
 32. BRAUER DJ, SCHMIDT BJ, PEARSON V. A framework for care during the stroke experience. *Rehabil Nurs* 2001 May-Jun;26(3):88-93.
 33. BURKE SO, KAUFFMANN E, COSTELLO E, WISKIN N, HARRISON MB. Stressors in families with a child with a chronic condition: an analysis of qualitative studies and a framework. *Can J Nurs Res* 1998 Spring;30(1):71-95.
 34. CARROLL SM. Nonvocal ventilated patients' perceptions of being understood. *West J Nurs Res* 2004 Feb;26(1):85-103.
 35. CHAPPLE A, ROGERS A. "Self-care" and its relevance to developing demand management strategies: a review of qualitative research. *Health Soc Care Community* 1999 Nov;7(6):445-54.
 36. CLEMMENS D. Adolescent motherhood: a meta-synthesis of qualitative studies. *MCN Am J Matern Child Nurs* 2003 Mar-Apr;28(2):93-9.
 37. COOK DJ, MEADE MO, PERRY AG. Qualitative studies on the patient's experience of weaning from mechanical ventilation. *Chest* 2001 Dec;120(6 suppl):469S-73S.
 38. DUGGAN F, BANWELL L. Constructing a model of effective information dissemination in a crisis. *Inf Res: An Int Electronic J* [serial online]. 2004 Apr;9(3):art. 178. [cited 4 Jan 2006]. <<http://informationr.net/ir/9-3/paper178.html>>.
 39. EVANS D, FITZGERALD M. The experience of physical restraint: a systematic review of qualitative research. *Contemp Nurse* 2002 Oct;13(2-3):126-35.
 40. FINFELD DL. Courage as a process of pushing beyond the struggle. *Qual Health Res* 1999 Nov;9(6):803-14.
 41. FINFELD DL. Self-resolution of drug and alcohol problems: a synthesis of qualitative findings. *J Addictions Nurs* 2000;12:65-72.
 42. FREDRIKSSON L. Modes of relating in a caring conversation: a research synthesis on presence, touch and listening. *J Adv Nurs* 1999 Nov;30(5):1167-76.
 43. FREDRIKSSON L, ERIKSSON K. The patient's narrative of suffering: a path to health? *Scand J Nurs Sci* 2001;1(1):s3-11.
 44. GARCIA J, BRICKER L, HENDERSON J, MARTIN MA, MUGFORD M, NIELSON J, ROBERTS T. Women's views of pregnancy ultrasound: a systematic review. *Birth* 2002 Dec;29(4):225-50.
 45. JONES ML. Application of systematic review methods to qualitative research: practical issues. *J Adv Nurs* 2004 Nov;48(3):271-8.
 46. KEARNEY MH. Enduring love: a grounded formal theory

- of women's experience of domestic violence. *Res Nurs Health* 2001 Aug;24(4):270-82.
47. LEFLER LL, BONDY KN. Women's delay in seeking treatment with myocardial infarction: a meta-synthesis. *J Cardiovasc Nurs* 2004 Jul-Aug;19(4):251-68.
 48. LEMMER B, GRELLIER R, STEVEN J. Systematic review of non-random and qualitative research literature: exploring and uncovering an evidence base for health visiting and decision making. *Qual Health Res* 1999;9:315-28.
 49. MCEWAN MJ, ESPIE CA, METCALFE J. A systematic review of the contribution of qualitative research to the study of quality of life in children and adolescents with epilepsy. *Seizure* 2004 Jan;13(1):3-14.
 50. MCKEVITT C, REDFERN J, MOLD F, WOLFE C. Qualitative studies of stroke: a systematic review. *Stroke* 2004 Jun;35(6):1499-505.
 51. MCNAUGHTON DB. A synthesis of qualitative home visiting research. *Public Health Nurs* 2000 Nov-Dec;17(6):405-14.
 52. MOLD F, MCKEVITT C, WOLFE C. A review and commentary of the social factors which influence stroke care: issues of inequality in qualitative literature. *Health Soc Care Community* 2003 Sep;11(5):405-14.
 53. MURRAY J, ASHWORTH R, FORSTER A, YOUNG J. Developing a primary care-based stroke service: a review of the qualitative literature. *Br J Gen Pract* 2003 Feb;53(487):137-42.
 54. NEILL S. Acute childhood illness at home: the parents' perspective. *J Adv Nurs* 2000 Apr;31(4):821-32.
 55. NELSON AM. A metasynthesis: mothering other-than-normal children. *Qual Health Res* 2002 Apr;12(4):515-30.
 56. NELSON AM. Transition to motherhood. *J Obstet Gynecol Neonatal Nurs* 2003 Jul-Aug;32(4):465-77.
 57. PATERSON BL. The shifting perspectives model of chronic illness. *J Nurs Scholarsh* 2001;33(1):21-6.
 58. ROBERTS KA, DIXON-WOODS M, FITZPATRICK R, ABRAMS KR, JONES DR. Factors affecting uptake of childhood immunisation: a Bayesian synthesis of qualitative and quantitative evidence. *Lancet* 2002 Nov 16;360(9345):1596-9.
 59. ROGERS WA. A systematic review of empirical research into ethics in general practice. *Br J Gen Pract* 1997 Nov;47(424):733-7.
 60. ROWE D, RUDKIN A. A systematic review of the qualitative evidence for the use of lifestyle planning in people with learning disabilities. *J Learning Disabil Nurs Health Soc Care* 1999;3(3):148-58.
 61. SANDELOWSKI M, BARROSO J. Toward a metasynthesis of qualitative findings on motherhood in HIV-positive women. *Res Nurs Health* 2003 Apr;26(2):153-70.
 62. SANDELOWSKI M, BARROSO J. Motherhood in the context of maternal HIV infection. *Res Nurs Health* 2003 Dec;26(6):470-82.
 63. SYDES MR, SPIEGELHALTER DJ, ALTMAN DG, BABIKER AB, PARMAR MKB, DAMOCLES Group. Systematic qualitative review of the literature on data monitoring committees for randomized controlled trials. *Clinical Trials* 2004 Feb;1(1):60-79.
 64. THORNE S, PATERSON B. Shifting images of chronic illness. *Image J Nurs Sch* 1998;30(2):173-8.
 65. THORNE S, PATERSON B, ACORN S, CANAM C, JOACHIM G, JILLINGS C. Chronic illness experience: insights from a metastudy. *Qual Health Res* 2002 Apr;12(4):437-52.
 66. WALTER FM, EMERY J, BRAITHWAITE D, MARTEAU TM. Lay understanding of familial risk of common chronic diseases: a systematic review and synthesis of qualitative research. *Ann Fam Med* 2004 Nov-Dec;2(6):583-94.
 67. WOODWARD V, WEBB C. Women's anxieties surrounding breast disorders: a systematic review of the literature. *J Adv Nurs* 2001 Jan;33(1):29-41.
 68. ARMAN M, REHNSFELDT A. The hidden suffering among breast cancer patients: a qualitative metasynthesis. *Qual Health Res* 2003 Apr;13(4):510-27.
 69. BARROSO J, POWELL-COPE GM. Metasynthesis of qualitative research on living with HIV infection. *Qual Health Res* 2000 May;10(3):340-53.
 70. BARROSO J, SANDELOWSKI M. Substance abuse in HIV-positive women. *J Assoc Nurses AIDS Care* 2004 Sep-Oct;15(5):48-59.
 71. BECK CT. Mothering multiples: a meta-synthesis of qualitative research. *MCN Am J Matern Child Nurs* 2002 Jul-Aug;27(4):214-21.
 72. BLEICH MR, HEWLETT PO, SANTOS SR, RICE RB, COX KS, RICHMEIER S. Analysis of the nursing workforce crisis: a call to action. *Am J Nurs* 2003 Apr;103(4):66-74.
 73. BRITTEN N, CAMPBELL R, POPE C, DONOVAN J, MORGAN M, PILL R. Using meta ethnography to synthesise qualitative research: a worked example. *J Health Serv Res Policy* 2002 Oct;7(4):209-15.
 74. CLARK AM, CURZIO J, LINDSAY GM, FLEMING VEM, MCINTOSH J. Exploring patients' perspectives of coronary heart disease: discerning methods and a review of the qualitative literature. *Coronary Health Care* 1998;2(3):118-28.
 75. HARDEN A, GARCIA J, OLIVER S, REES R, SHEPHERD J, BRUNTON G, OAKLEY A. Applying systematic review methods to studies of people's views: an example from public health research. *J Epidemiol Community Health* 2004 Sep;58(9):794-800.
 76. JENSEN LA, ALLEN MN. A synthesis of qualitative research on wellness-illness. *Qual Health Res* 1994;4:349-69.
 77. KEARNEY MH. Truthful self-nurturing: a grounded formal theory of women's addiction recovery. *Qual Health Res* 1998 Jul;8(4):495-512.
 78. KENNEDY HP, ROUSSEAU AL, LOW LK. An exploratory metasynthesis of midwifery practice in the United States. *Midwifery* 2003 Sep;19(3):203-14.
 79. MCCORMICK J, RODNEY P, VARCOE C. Reinterpretations across studies: an approach to meta-analysis. *Qual Health Res* 2003 Sep;13(7):957-73.
 80. MEADOWS-OLIVER M. Mothering in public: a meta-synthesis of homeless women with children living in shelters. *J Spec Pediatr Nurs* 2003 Oct-Dec;8(4):130-6.
 81. MORSE JM. Responding to threats to integrity of self. *ANS Adv Nurs Sci* 1997 Jun;19(4):21-36.
 82. RUSSELL CK, BUNTING SM, GREGORY DM. Protective care-receiving: the active role of care-recipients. *J Adv Nurs* 1997 Mar;25(3):532-40.
 83. SANDELOWSKI M, BARROSO J. Finding the findings in qualitative studies. *J Nurs Scholarsh* 2002;34(3):213-9.
 84. SANDELOWSKI M, BARROSO J. Creating metasummaries of qualitative findings. *Nurs Res* 2003 Jul-Aug;52(4):226-33.
 85. SANDELOWSKI M, LAMBE C, BARROSO J. Stigma in HIV-positive women. *J Nurs Scholarsh* 2004;36(2):122-8.
 86. SHERWOOD G. Meta-synthesis of qualitative analyses of caring: defining a therapeutic model of nursing. *Adv Pract Nurs Q* 1997 Summer;3(1):32-42.
 87. STAVRI PZ. Personal health information-seeking: a qualitative review of the literature. *Stud Health Technol Informatic* 2001;84(2):1484-8.
 88. VARCOE C, RODNEY P, MCCORMICK J. Health care relationships in context: an analysis of three ethnographies. *Qual Health Res* 2003 Sep;13(7):957-73.
 89. WARR DJ. Stories in the flesh and voices in the head: reflections on the context and impact of research with dis-

advantaged populations. *Qual Health Res* 2004 Apr;14(4): 578–87.

90. MOHER D, COOK DJ, EASTWOOD S, OLKIN I, RENNIE D, STROUP DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. quality of reporting of meta-analyses. *Lancet* 1999 Nov 27; 354(9193):1896–900.

91. STROUP DF, BERLIN JA, MORTON SC, OLKIN I, WILLIAMSON GD, RENNIE D, MOHER D, BECKER BJ, SIPE TA, THACKER SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. *JAMA* 2000 Apr 19;283(15):2008–12.

92. GOLDER S, MATHER L, WRIGHT K, GLANVILLE J. What's on the Web: recommendations for reporting research. *J Health Serv Res Policy* 2005;10(2):124–5.

93. MULROW CD. The medical review article: state of the science. *Ann Intern Med* 1987 Mar;106(3):485–8.

94. MCALISTER FA, CLARK HD, VAN WALRAVEN C, STRAUS SE, LAWSON FM, MOHER D, MULROW CD. The medical review article revisited: has the science improved? *Ann Intern Med* 1999 Dec 21;131(12):947–51.

95. PRITCHARD SJ, WEIGHTMAN AL. Towards a Cochrane Information Retrieval Methods Group: a progress report. *Health Info Libr J* 2003 Jun;(20 suppl 1):69–71.

Received September 2005; accepted April 2006

Table 2

Other qualitative systematic reviews excluded for not reporting search methods

Study	Reason for exclusion*
Arman (2003) [68]	No details
Barroso and Powell-Cope (2000) [69]	No details, mentions "Extensive computer searches"
Barroso and Sandelowski (2004) [70]	No details (reported in [27])
Beck (2003) [30]	No details (reported in [71])
Bleich et al. (2003) [72]	No details
Britten et al. (2002) [73]	No details, studies arbitrarily chosen from previous review
Clark et al. (1998) [74]	No details
Harden et al. (2003) [75]	No details
Jensen and Allen (1994) [76]	No details
Kearney (1998) [77]	No details
Kennedy et al. (2003) [78]	Only included studies conducted by authors
McCormick et al. (2003) [79]	Only included studies conducted by authors
Meadows-Oliver (2003) [80]	No details
Morse (1997) [81]	No details
Russell et al. (1997) [82]	Only included studies conducted by authors
Sandelowski and Barroso (2002) [83]	No details
Sandelowski and Barroso (2003) [84]	No details (reported in [27])
Sandelowski et al. (2004) [85]	No details (reported in [27])
Sherwood (1997) [86]	No details
Stavri (2001) [87]	No details
Varcoe et al. (2003) [88]	Only included studies conducted by the authors

* "No details" signifies no details of *either* databases *or* keywords *or* other methods.**Table 4**

Outline example of a report of a literature search structured according to STARLITE principles

Title: Qualitative systematic review of support for breastfeeding [Publication pending]	
S: Sampling strategy	Purposive: Samples two databases from medicine, nursing, and social science fields
T: Type of study	Any kind of qualitative study (includes ethnographic, grounded theory, focus groups, etc.)
A: Approaches	Subject searching, citation searching, hand-searching, Internet searching, contact with experts
R: Range of years (start date: end date)	No restrictions: to the beginning of each candidate database—to the end of 2003
L: Limits	English, human
I: Inclusion and exclusions	Inclusion: qualitative method, about support for breast-feeding women; exclusion: quantitative method, animal lactation, physiology of breast milk (i.e., not related to aspects of support)
T: Terms used	Complete search strategies available as additional files via Website
E: Electronic sources	MEDLINE [Ovid], EMBASE [Ovid], CINAHL [Ovid], British Nursing Index [SilverPlatter], Applied Social Sciences Index and Abstracts (ASSIA) [Cambridge Scientific Abstract], Social Sciences Citation Index [Web of Knowledge], other databases (e.g., MIDIRS, NeLH)

Appendix C - Paper M2

This is the Post-print version with changes from referees' comments.

Publisher copyright and the source are acknowledged as follows:

Dixon-Woods, M, Booth, A & Sutton AJ. Synthesizing qualitative research: a review of published reports *Qualitative Research* 2007 Aug 7: 375-422

SYNTHESISING QUALITATIVE RESEARCH: A REVIEW OF PUBLISHED REPORTS

Mary Dixon-Woods,¹ Andrew Booth² Alex J Sutton³

¹ Reader in Social Science and Health, Department of Health Sciences, University of Leicester, 22-28 Princess Road West, Leicester, LE1 6TP, UK. Email: md11@le.ac.uk
Fax: 0116 2523204 Phone: 0116 2523272 (corresponding author)

² Director of Information Resources and Reader in Evidence-based Information Practice, School of Health And Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK.

³ Senior Lecturer in Medical Statistics, Department of Health Sciences, University of Leicester, 22-28 Princess Road West, Leicester LE1 6TP, UK.

SYNTHESISING QUALITATIVE RESEARCH: A REVIEW OF PUBLISHED REPORTS

Abstract

Although there is increasing demand for syntheses of qualitative research, little is known about papers that aim to report such syntheses. We searched for published reports of attempts to conduct syntheses of qualitative research in health and healthcare. Papers were included if they were published between 1988 and 2004, in the English language, and in a peer-reviewed journal. We identified a modest body of literature (42 papers) reporting syntheses of qualitative research in health and healthcare. We extracted data on the topic of the paper and reported methods for searching, appraisal, and synthesis. Many papers lack explicitness about methods for searching, appraisal, and synthesis, and there is little evidence of emerging consensus on many issues. Some papers reported purposive attempts to innovate with, and to adapt, methods for synthesis. There was also some evidence of possibly inappropriate use of some techniques. We conclude that continued methodological progress and improved reporting are required.

Keywords: Qualitative synthesis, systematic reviews, meta-synthesis, meta-ethnography, evidence-based policy and practice, searching techniques

Background

Systematic review has developed as a specific methodology for research synthesis involving explicit methods of searching for, appraising, and summarising and combining evidence. It has rapidly become a cornerstone of the evidence-based practice and policy movement (Dixon-Woods et al, 2006a). Systematic reviews, as conventionally understood, have specific characteristics: an explicit study protocol, addressing a pre-specified, highly focused question(s); explicit methods for searching for studies; appraisal of studies to determine their scientific quality; and explicit methods for combining the findings across a range of studies (Egger et al, 1995). The advantages of systematic review methodology are seen to lie in its rigour and transparency of process, and the avoidance of the “biases” associated with informal or inexplicit methods for identifying and assessing evidence (Mulrow, 1987; McAlister et al, 1999; Klassen, Jadad and Moher, 1998).

Although its clearest impact to date is seen in the area of health and healthcare, systematic review is rapidly migrating to other areas, including education and social policy (<http://www.campbellcollaboration.org/>). However, the limitations of current methods for systematic review, which tend to focus on questions of effectiveness, and which tend to privilege quantitative forms of evidence, have become evident (Mays, Pope, Popay, 2005; Greenhalgh et al, 2005). Increasingly, practitioners, policy-makers and managers demand syntheses of evidence that acknowledge complexity and context (Bravata et al, 2005). There has been a corresponding recognition that excluding qualitative evidence from systematic reviews may neglect useful information (Sheldon, 2005) or even distort the evidence-base. Incorporating qualitative research in systematic reviews poses daunting methodological problems, however (Dixon-Woods, Fitzpatrick, Roberts, 2001), and recent years have seen a rapid expansion of activity devoted to developing and critiquing methods for conducting syntheses of qualitative research (Dixon-Woods et al, 2005).

Although the number of publications reporting syntheses of qualitative research is now rapidly increasing, little is known about which methods for synthesis are used and with what

frequency, and how such syntheses deal with key challenges of review methodology, including methods for searching and appraisal. These are important questions, particularly in contributing to debates about how far principles of (conventional) systematic review have been incorporated into methodological practices in this newer area (Dixon-Woods et al, 2006a). We aimed to conduct a structured review of published reports of attempts to conduct syntheses of qualitative research in health and healthcare.

Methods

To be included in our review, a paper had to meet all of the following criteria:

- Conducted within a health or healthcare context
- Reporting a synthesis of qualitative research (or a synthesis of qualitative with quantitative research) by synthesis methods other than informal review. Papers commenting on methodological issues but without including details of the outcomes of the synthesis were excluded, as were publications that did not explicitly describe or name a method for synthesis.
- Published between 1988 (the date of publication of Noblit and Hare's key text on qualitative synthesis (Noblit, & Hare, 1988)) and 2004
- Published in the English language
- Published in a peer-reviewed journal.

In specifying these inclusion criteria, we did not intend them to function as standards, only as means by which the account of the papers could be standardised around a common set of issues to allow a focused discussion and reflection. The choice of these specific criteria was guided by our interest in assessing how far current practices in synthesising qualitative research are similar to or different from those used in conventional systematic review, though we did not have a normative commitment to the "superiority" of any particular approach.

As reported elsewhere (Booth, in press) a comprehensive search was undertaken of the social science, health and information science literature on PubMed, MEDLINE, CINAHL, Web of Knowledge (including the Science and Social Sciences Citation Index) and the Cochrane Methodology Register. A sensitive search strategy was used, combining the keywords "(qualitative review\$) OR metasynthesis OR meta-synthesis OR (qualitative AND systematic review\$)". Citations to key qualitative methodological texts were followed up (Noblit, & Hare, 1988; Paterson, Thorne, Canam & Jillings, 2001), and the "Related Articles" features on Medline and Web of Science were utilised.

Each paper was examined in detail under the following categories:

- Topic of review.
- Strategies for searching, including databases searched, keywords used, any other search strategies, and the extent to which searches aimed to be comprehensive. Comprehensiveness referred to how far the authors aimed to include all relevant papers in the area of their synthesis, compared with (for example) using a sample.
- Number of publications included in the review.
- Methods used for appraising the quality of publications included in the review.
- Methods used for synthesising publications included in the review.

In some cases authors referred to their previous papers for methodological details and did not provide a description of methods of searching, appraisal, and so on in the paper at hand. Unless details were specified in the paper at hand, they were judged not to have been explicit.

Results

Searching identified in excess of 400 articles that passed an initial screening by AB on title. Further assessment by two reviewers (AB, MDW) using detailed review of abstracts and full text (where necessary) from those articles yielded 42 publications that met the inclusion

criteria. Several studies described searching for and reviewing or synthesising qualitative studies, without explicitly describing a method for synthesis: examples included a review of self-care (Chapple, Rogers, 1999) and a review of patients' experiences of being weaned from mechanical ventilation (Cook, Meade & Perry, 2001). These were excluded. Table 1 summarises the data extracted from each paper included in our review.

Characteristics of papers

We did not identify any papers meeting the inclusion criteria that were published before 1994. Ten papers were published between 1994 and 1999, while 33 were published between 2000 and 2004, with 12 published in 2003 and seven in 2004. The majority were published in nursing journals (25/43) or concerned nursing-related topics. One journal (*Qualitative Health Research*) published almost a fifth (8/42) of the papers. Of the 42 included papers, 30 were published in North American journals. A small number of authors appear to be active in the area, with some authors making repeated (up to five) contributions to the dataset, but the majority contributing only once.

The number of studies synthesised in the included papers ranged from three to 292 (median 15). Five papers described themselves as a "systematic review" in their title. Some studies explicitly identified themselves as undertaking methodological research on particular elements of review methodology. For example *Papers 9 and 11* presented worked examples of meta-ethnography, while *Paper 33* presented a demonstration study of Bayesian meta-analysis.

Methods for searching

Sixteen papers did not specify any methods for searching, and five papers described using the authors' own papers as either part or whole of the sample. The databases that were searched to identify candidate studies for inclusion in reviews were specified by 27 papers. The database most frequently mentioned was CINAHL (22) closely followed by Medline (21). Most common supplementary strategies used alongside bibliographic databases were following up reference lists (eight times) and hand-searching (five times). The most

comprehensive report of the literature searching process, *Paper 8*, itemised 23 databases, 10 websites, six other techniques and also reported the keywords (search terms).

Keywords used for searching were reported in fewer than half (19) of the papers. We have not reported full details of keyword terms because of lack of space, but many keyword search strategies revealed a lack of sophistication, including a lack of clarity regarding use of free text terms or approved subject headings. A common limitation of the search strategies reported was a failure to use methodological terms in addition to subject-specific terms (e.g. the term “qualitative”) of the type that would increase specificity in retrieving qualitative studies.

Methods of appraisal

Twenty-one papers did not describe appraisal of candidate studies for their reviews, nor did they offer a justification for not appraising studies. Six papers (18, 20, 33, 35, 36 and 37) explicitly described not conducting formal appraisals of studies, usually because they did not use study quality as an exclusion criterion. Five papers (10, 21, 22, 30, and 32) offered an account of how they assessed quality of papers that did not involve using a formal checklist. Seven papers (2,3,11,16,28,38,38, 39 40 and 42) described modifying existing instruments or criteria to appraise studies; only one used an existing instrument without modification.

It was not uncommon for papers to describe methods of appraisal in detail, but then to fail to give an account of whether the judgements of quality were used to exclude papers from the review, or how the outcomes of the appraisals were taken into account in the synthesis.

Methods of synthesis

By far the most commonly used method of synthesis was meta-ethnography, cited by 19 papers. Other methods of synthesis included meta-synthesis/metasummary/meta-study (9); meta-interpretation/interpretive synthesis (3); the constant comparative method/grounded theory (3); qualitative meta-analysis (2); content analysis (2); interpretative synthesis based

on the hermeneutic circle (1); hermeneutic phenomenology(1); Miles and Huberman's principles of cross-case analysis (1) and clustering (1).

In some cases authors have explicitly and purposefully innovated with the methods of meta-ethnography. *Paper 11*, for example, built on Schutz's (1962) distinction between first order constructs – the “constructs of everyday life” and second order constructs, which are the constructs used by social scientists. *Paper 11* proposed that the products of a meta-ethnography can be deemed to be third order constructs, as they are interpretations of social science constructs. Constructs reported in the original studies would therefore be deemed to be second order constructs. In *Paper 42*, however, it was not clear that this approach was directly applied: the authors describe their synthesis as involving “determining the key concepts from each article, known as the first-order constructs; translating the first-order constructs across articles to determine second-order constructs”, suggesting possible lack of clarity about the distinction between first and second order constructs.

Some papers, although reporting use of techniques such as meta-ethnography, appeared either to be reporting findings that were not recognisable as a meta-ethnography, or to be deviating from recognised principles and procedures. *Paper 20*, for example, despite citing Noblit and Hare, do not appear to report an analysis that is recognisably a meta-ethnography. In several cases (eg *paper 40*) it is not clear how the findings of a review conducted using meta-ethnography might be distinct from those using narrative review.

Discussion

Although there is growing interest in how qualitative research can be synthesised for purposes of informing policy and practice, we have identified only a modest body of literature published between 1988 and 2004 that has used explicit methods for synthesis. It is clear, however, that synthesis of qualitative research is rapidly expanding and our sample suggests it is now beginning to migrate from nursing, where it has traditionally had its strongest base, to other areas of health and social care. Meta-ethnography is by far the most commonly used synthesis technique, but other methods are now also being developed and used. Some existing work is explicitly aimed at developing and evaluating methods for synthesis, and has

therefore focused on particular elements of review methodology. Much of the applied work, however, lacks explicitness about methods for searching, appraisal, and synthesis, and there is little evidence of an emerging consensus on issues such as methods for appraisal of included studies.

Our review has a number of limitations. Some of these relate to the construction of the sample. As with primary qualitative research, (Shaw et al, 2004) there are considerable difficulties in searching for reviews and syntheses of qualitative research, in part due to variations in the terminology used to describe such reviews. This creates “false-negatives” (publications that fail to be identified as reviews of qualitative research) as well as “false positives” (some reviews appear to be “qualitative systematic reviews” but are simply using the term “qualitative” to indicate that they have not used a formal method of quantitative synthesis such as meta-analysis). Few of the journals in which reviews of qualitative literature are published utilise structured abstracts, which can help in identifying papers using relevant methods (Booth and O'Rourke, 1997).

Notwithstanding these limitations, our findings raise interesting and important questions about the way forward in methods for synthesis of qualitative research. Our findings suggest that many papers reporting syntheses of qualitative research lack explicit description of at least some methods. Many do not explain how they identified their sample, whether or how they appraised included papers, or how appraisal judgements were accounted for in the synthesis. Many papers in our review offered no defence of their lack of explicitness in describing their techniques of searching; nearly 40% simply did not describe how the studies were identified at all (though a small number of papers defended the selection of a particular sample on grounds that the aim of the work was methodological development). Many did not explain whether they were aiming for comprehensiveness or were using some form of sampling strategy.

We are not arguing that reviews of qualitative studies must necessarily use precisely the kinds of structured searching techniques that have become standard practice within conventional systematic review: it may well be that some of the practices used in papers in our sample are indeed defensible and desirable. For example, in relation to searching, the

limitations of relying exclusively on protocol-driven approaches to searching for evidence in complex areas have recently been identified (Greenhalgh and Peacock, 2005). Similarly, the limitations of aiming for “comprehensiveness” have also recently recognised, and alternative approaches such as using principles of theoretical sampling and theoretical saturation have been proposed as a means of constructing the boundaries of reviews (Dixon-Woods et al, 2006b). But without theoretical and methodological justifications such as these of how the studies in a review were selected for inclusion, criticisms that the choice of studies is idiosyncratic and capricious, and likely to offer a partial and potentially misleading account of the evidence, are likely to persist (Chalmers et al, 2002)

Also troubling are suggestions from our sample that even where papers are explicit about what they have done, the methods are not always applied optimally. For example, even where searching techniques were described, these techniques were not always rigorous, failing to make best use of searching strategies that are known to improve precision and yield (Shaw et al, 2004). Searching techniques were often informal or lacking in sophistication, and were likely to miss relevant material. Although the benefits of including an information specialist skilled in searching and retrieval techniques on review teams have long been recognised in the field of quantitative syntheses (Smith, 1996), few reviews in our sample appeared to include such an individual. Again, it is important that appeals to the particularistic nature of qualitative research, however valid, are not used to legitimate or obscure what may in fact be poor practices and faults of execution.

The finding that there was little uniformity in the choice of methods for appraisal was perhaps unsurprising, given on-going disagreement among practitioners not only about the characteristics that define good quality qualitative research, but also on whether criteria for quality in qualitative research should exist at all (Dixon-Woods et al, 2004). Attempts to resolve the impasse have proved difficult; most efforts seem to result in yet another list of quality criteria without generating any greater consensus. Indeed, how far criteria-based approaches offer significant advantages over expert intuitive judgement in assessing the quality of qualitative research is being challenged by recent evidence indicating that checklist-style approaches may be no better at promoting agreement between reviewers (Dixon-Woods

et al, in press). What is perhaps rather more disappointing in our sample was the failure in most papers we reviewed to incorporate judgements of quality in any explicit way in the synthesis; for some syntheses at least, the quality appraisal appeared to be simply a procedural step with little significance for the eventual conclusions. It could indeed be argued that “critical appraisals” of the type used in quantitative syntheses are less appropriate for reviews of qualitative evidence, where the purpose of the review is more likely to be oriented towards maximising the conceptual yield of included papers rather than determining the robustness of the study design so that sensitivity analyses can be conducted. Rather than appraisals of the procedural defects of individual papers, recent work is beginning to suggest instead a turn towards a more critique-led approach that takes the corpus of a body of literature as its object of inquiry (Greenhalgh et al, 2005; Dixon-Woods et al, 2006b).

Our analysis identified that some methods for synthesis are used much more frequently than others. As the number of available techniques for synthesis continues to increase, more diversity can be anticipated. Of particular note was the tendency in the papers in our review to borrow existing methods, especially for synthesis, and to make significant adaptations to these. This is clearly not a problem where the aim is methodological development and where authors are explicit about describing and demonstrating the nature of their methodological innovations. However, making amendments to methods without being explicit, and making significant amendments while still retaining the label of the original method, is rich in potential for confusion. Thorne et al (2004) comment on this issue, criticising an apparent trend for authors to claim to have used the “metasynthesis” techniques developed by their group to “repackage” or legitimate ordinary literature reviews. It is indeed unhelpful if, through incremental adaptation, a technique such as meta-ethnography, exists in many variants, so that it is no longer clear what is meant when the term is used. Even more troubling is evidence that recognised methodologies are sometimes interpreted or used inappropriately. The problems that such practices have caused in the area of grounded theory are well recognised (Pidgeon and Henwood, 2004).

Clearly, many papers in our review that are reporting qualitative syntheses would not fulfil the conventional criteria for systematic review. Because of the problems in reporting of the

syntheses that we have identified, it is at present difficult to distinguish whether this is because of genuine differences of methodological principle between systematic review and synthesising qualitative research (which are likely to be defensible) or differences in explicitness and quality of execution (which are likely to be less defensible). For example, it may well be preferable for syntheses of qualitative research to use methods of searching, appraisal, and synthesis that are quite distinct from those used in (conventional) systematic review; what is less acceptable is when these methods are not adequately justified or are poorly undertaken or reported. If syntheses of qualitative research are to gain credibility, they must use methods in ways that are methodologically defensible and provide explicit accounts of these.

One possible way of improving the clarity of reporting of syntheses of qualitative research might involve the development of a set of reporting guidelines similar to those developed for meta-analyses of trials (Moher et al, 1999 & Shea et al, 2000) , epidemiological studies (Stroup et al, 2000) and studies of animal experiments (Peters et al, in press). Where such guidelines have been used to critique the conduct and reporting of the quantitative synthesis literature (Christensen, 2001; Hemels et al, 2004 & Peters et al, in press), they have found evidence of significant variations in reporting standards. A move towards improved explicitness about reporting of syntheses of qualitative research could take place ahead of a consensus emerging on methods for synthesis, and would likely contribute to improved reflexivity and better research practice (Elliott, 2005).

Conclusions

Reviews of complex bodies of evidence that incorporate qualitative evidence are increasingly sought. Our review of articles that have attempted to conduct syntheses of qualitative research suggests that, although the area is developing rapidly, there are many methodological challenges and standards of practice that must be confronted. It will be helpful if future syntheses are explicit and reflexive, and if they document their methods carefully.

Acknowledgements: We would like to thank the ESRC Research Methods Programme (Award Number A333250043) for funding the work on which this paper is based. We would like to thank our colleagues on this award (David Jones, Tina Miller, Rachel Shaw, Jonathan Smith, and Bridget Young) and Professor Alan Bryman, who provided very helpful comments on an earlier draft of this paper.

Conflicts of interest: The authors of this paper are also authors of two of the papers included in the review (MDW - paper 33; AB – paper 8).

Table 1: Characteristics of papers included in review

	Paper	Review subject	Methods for searching	Number of papers included in review [of papers initially identified]	Methods of appraisal of papers	Methods of synthesis
1. 1	Arman M, Rehnsfeldt A. The hidden suffering among breast cancer patients: a qualitative meta-synthesis. Qualitative Health Research 2003; 13: 510-527	Lived experience of breast cancer/suffering	Databases: Not specified Keywords: Specified Other search strategies: Not described Comprehensiveness: Unclear. Limited to scientific articles written in English, published in nursing/caring journal 1990-200, using a qualitative or qualitative/quantitative method.	14 [no denominator]	Not described	Hermeneutic phenomenological approach
2.	Attree P. Growing up in disadvantage: a systematic review of the qualitative evidence. Child: Care, Health & Development. 2004; 30:679-699	Children's perspectives on growing up in disadvantage	Databases: Not specified Keywords: Not specified Other search strategies: "Grey" literature sources, reference lists, websites, paper journals, key informants. Comprehensiveness: Comprehensive. English only. Date restricted.	9/61 [11,224]	Modified checklist based on previous approaches, including 10 questions on research background; aims and objectives; study context; appropriateness of design; sampling; data collection; data analysis; reflexivity; contribution to	Meta-ethnography

					knowledge; research ethics. Quality appraisal carried out by two reviewers independently and papers graded as A (no or few flaws) to D (seriously flawed). Only studies rated A or B included.	
3.	Barroso J, Powell Cope GM. Meta-synthesis of qualitative research on living with HIV infection. Qualitative Health Research. 2000; 10: 340-353	Living with HIV infection	<p>Databases: Not specified</p> <p>Keywords: Not specified</p> <p>Other search strategies: Not described</p> <p>Comprehensiveness: Not clear. Studies limited to those published in refereed journals, reporting on people living in the USA. English only. Date restricted.</p>	21/45	Developed and pilot tested a detailed critique form based on Burns' standards for qualitative research: descriptive vividness, methodological congruence, analytical preciseness, theoretical connectedness, heuristic relevance. Articles receiving unfavourable ratings from two reviewers were excluded.	Constant comparative method
4.	Barroso J, Sandelowski M. Substance abuse in HIV positive women. Journal of the Association of Nurses in AIDS care 2004; 15:48-59.	Substance abuse in HIV-positive women	<p>Databases: Not specified</p> <p>Keywords: Not specified</p> <p>Other search strategies: Not specified</p> <p>Comprehensiveness: Comprehensive</p>	74/114	Not described	Meta-summary, meta-synthesis

5.	Beck CT. Caring within nursing education: a metasynthesis. Journal of Nursing Education. 2001; 40: 101-109	Caring within nursing education	<p>Databases: CINAHL</p> <p>Keywords: Specified</p> <p>Other search strategies: No details</p> <p>Comprehensiveness: Comprehensive. English only. Date restricted.</p>	14/n.d.	Not described	Meta-ethnography
6.	Beck CT. Mothering multiples: a meta-synthesis of qualitative research. MCN. 2002; 27: 214-221	Mothering multiples in the first year of life	<p>Databases: CINAHL, Psylit (sic), and Medline</p> <p>Keywords: Not specified</p> <p>Other search strategies: Not described</p> <p>Comprehensiveness: Comprehensive. English only. Date restricted.</p>	6/n.d.	Not described	Meta-ethnography
7.	Beck CT. Postpartum depression: a metasynthesis. Qualitative Health Research. 2002; 12: 453-472	Postpartum depression	<p>Databases: Cumulative Index to Nursing and Allied Health Literature (CINAHL), Psychlit, Index Medicus, Social Sciences Citation Index, American Humanities Index, Anthropological Literature, Sociological Abstracts. 1960s-1990s.</p> <p>Keywords: Specified</p> <p>Other search strategies: Unpublished studies through Dissertation Abstracts.</p>	18/n.d.	Not described	Meta-ethnography

			Comprehensiveness: Comprehensive. English only. Date restricted			
8.	Beverley CA, Bath PA, Booth A. Health information needs of visually impaired people: a systematic review. Health and Social Care in the Community. 2003; 12: 1-24	Health information needs of visually impaired people	Databases: Allied and Alternative Medicine Database (AMED); • Applied Social Sciences Index and Abstracts (ASSIA); British Education Index; British Nursing Index; Caredata; Cochrane Controlled Trials Register (CCTR); Cochrane Database of Systematic Reviews (CDSR); Cumulative Index to Nursing and Allied Health (Cinahl); Database of Abstracts of Reviews of Effectiveness (DARE); Department of Health Library Database (DH-DATA); Educational Resources Information Center (ERIC); Embase; Health Technology Assessment (HTA); Helmis; INSPEC; International Bibliography of the Social Sciences (IBSS); King's Fund Database; Library and Information Science Abstracts (LISA); MEDLINE; PreMEDLINE; PsycINFO; Science Citation Index; Social Sciences Citation Index; 'Grey literature' databases: Current Research in Britain (CRiB); Health Development Agency (HDA) Evidence Base;	16/114	Included studies were critically appraised using the Critical Skills Training in Appraisal for Librarians checklist	Meta-ethnography

			<p>Health Management Information Consortium (HMIC); Index to Theses; National Research Register (NRR); UK Official Publications (UKOP);</p> <p>Keywords: Specified</p> <p>Other search strategies: 28 Websites/search engines. Citation searching of Science Citation Index and, Social Science Citation Index; 12 Organisations contacted 'Grey literature' sources were searched, Websites of relevant organisations and research registers supplemented by general Internet search using meta-search engine (Copernic). Major organisations and experts contacted. Reference lists checked.</p> <p>Comprehensiveness: Comprehensive. English only. Date restricted.</p>			
9.	<p>Britten N, Campbell R, Pope C, Donovan J, Morgan M, Pill R. Using meta-ethnography to synthesise qualitative research: a worked example. Journal of Health Services Research and Policy. 2002; 7:209-15.</p>	<p>Lay meanings of medicines</p>	<p>Databases: Not specified</p> <p>Keywords: Not specified</p> <p>Other search strategies: Studies chosen arbitrarily, including three by the authors</p> <p>Comprehensiveness: Sampling was purposive</p>	4/n.d.	Not described	Meta-ethnography

10.	Burke SO, Kauffmann E, Costello E, Wiskin N, Harrison MB. Stressors in families with a child with a chronic condition: an analysis of qualitative studies and a framework. Canadian Journal of Nursing Research. 1998; 30:71-95.	Stressors in families with a child with a chronic illness	Databases: CINAHL and MEDLINE Keywords: Specified Other search strategies: Studies from a previous review, pre-1990. Comprehensiveness: Comprehensive. English only. Date restricted	17/n.d.	Only studies that clearly described the subjects were included. Enough description to infer [sic] use of a qualitative approach had to be provided. Trustworthiness of results was assessed.	Adapted from qualitative "meta-analysis" methods described by Morse and Johnson (1991), Noblit and Hare (1988) and Thorne (1994).
11.	Campbell R, Pound P, Pope C, Britten N, Pill R, Morgan M, Donovan J. Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. Social Science and Medicine. 2003; 56: 671-684	Lay experiences of diabetes and diabetes care	Databases: BIDS IBSS database Keywords: Specified Other search strategies: own collections of papers, hand searches of three journals: Qualitative Health Research (all years), Sociology of Health and Illness and Social Science and Medicine (previous 20 years). Comprehensiveness: Purposive: first ten papers meeting criteria. English only. Date restricted	7/10	Each paper was assessed by two reviewers using criteria based on the CASP tool, specially adapted by the authors for use in a synthesis.	Meta-ethnography
12.	Carroll, SM (2004) Nonvocal ventilated patients' perceptions of being understood. Western Journal of Nursing Research. 2004; 26: 85-103	Mechanically ventilated patients' experiences with communication.	Databases: CINAHL, MEDLINE, Sociofile, Psychlit, Dissertation Abstracts Keywords: Specified Other search strategies: No details	12/n.d.	Not described	Meta-synthesis, including synthesis of data, methods and theoretical frameworks of the studies

			Comprehensiveness: Comprehensive. English only. Date restricted			
13.	Clemmens D. Adolescent motherhood: a meta-synthesis of qualitative studies. MCN 2003; 28: 93-99	Adolescent motherhood	Databases: CINHALL (sic), Medline, PsychINFO (sic), ERIC, Sociological Abstracts, and Dissertation Abstracts. Keywords: Specified Other search strategies: None Comprehensiveness: Purposive: Studies conducted in United States, Canada, China, England, and Australia. English only. Date restricted.	25/50 [251]	Not described	Meta-ethnography
14.	Coffman MJ. Cultural caring in nursing practice: a metasynthesis of qualitative research. Journal of Cultural Diversity. 2004; 11: 100-109	Healthcare clinicians' experiences of caring for patients from other cultures	Databases: CINAHL, Medline, PsychINFO, ERIC, Sociological Abstracts and Dissertations Abstracts Online. Keywords: Specified Comprehensiveness: Comprehensive. Restricted to experiences of nurses only after initial searches.	13/n.d.	Not described.	Meta-ethnography
15.	Duggan F, Barnwell L. Constructing a model of effective information dissemination in a crisis. Information Research. 2004; 9. http://informationr.net/ir/9-	Dissemination of information to the community during a health crisis	Databases: ANTE, Arts and Humanities Index, ASSIA, ASTI, British Education Index, Cinahl, Dissertation Abstracts Online, ERIC, Information Science Abstracts, Library Literature, LISA, Medline, Psyclit, Social Sciences	20/n.d.	Not described	Meta-ethnography

	3/paper178.html		<p>Citation Index, Sociological Abstracts, Wilson Social Science Abstracts, BUBL, Ingenta Journals, Library Management, OMNI, SOSIG, Uncover, University Web sites, Update-software.com, BMJ customised alert service, Contents direct table of contents service, lis-medical mailing list, UNN library catalogue.</p> <p>Keywords: Not specified</p> <p>Other search strategies: Handsearch of most frequently occurring journal and references of relevant papers</p> <p>Comprehensiveness: Comprehensive. English only. Date restricted</p>			
16.	Evans D, FitzGerald, M. The experience of physical restraint: A systematic review of qualitative research Contemporary Nurse 2003; 13: 126-135	Experience of physical restraint from the perspectives of the person subject to restraint and of their family	<p>Databases: CINAHL, MEDLINE, Current Contents, PsycInfo, Embase, HealthSTAR and the Expanded Academic Index.</p> <p>Keywords: Specified</p> <p>Other search strategies: Reference lists of all identified reports and articles searched for additional studies.</p> <p>Comprehensiveness: Comprehensive. English only. Date restricted</p>	4/310 (4,500+)	A locally developed tool. Criteria were: appropriate methodology, clear description of both the population and study method, adequate result information supported by exemplars from study participants.	New interpretive synthesis method adapted from meta-ethnography, meta-synthesis, content analysis, and grounded theory methods.

17.	Finfgeld, DL. Courage as a process of pushing beyond the struggle. Qualitative Health Research 1999; 9: 803-814	Courage among persons who were experiencing a variety of lingering threats to their well-being	Databases: Medline, PsychInfo, and CINAHL Keywords: Not specified Other search strategies: Not specified Comprehensiveness: Comprehensive. English only. Date restricted	6/n.d.	Inclusion criteria specified that papers must : report data that was not collected incidentally and were not primarily anecdotal in nature; were conducted using pre-determined, explicit, and widely accepted qualitative methods; and provided adequate information to substantiate the latter criterion.	Meta-interpretation, based on meta-ethnography but argued to have broader connotations
18.	Frederiksson, L.; Eriksson, K. The patient's narrative of suffering: a path to health? Scandinavian Journal of Nursing Sciences, 2001. 1: s. 3 – 11.	Patients' narratives of suffering	Databases: CINAHL only Keywords: Specified Other search strategies: manual search of reference lists. Comprehensiveness: Comprehensive. In English or a Scandinavian language. Date restricted	23/4438	Critical evaluation was not undertaken before inclusion of papers in the review, but was completed instead as part of the synthesis process, where a weak study would stand out as an anomaly or because supporting data are missing.	Interpretative synthesis based on meta-ethnography
19.	Fredriksson L. Modes of relating in a caring conversation: a research synthesis on presence, touch and listening. Journal of Advanced Nursing 1999; 30: 1167-76.	Ontological and theoretical understanding of presence, touch, and listening in a caring conversation	Databases: CINAHL only Keywords: Specified Other search strategies: manual search of reference lists. Comprehensiveness: Purposive to assure inclusion of relevant	28/1681	Not described	Based on hermeneutic circle. Interpretative task is divided into three phases: isolation of parts of study which describe the phenomenon; find a common structure within which studies can be translated into

			studies only and to limit sample to manageable size. In English or a Scandinavian language. Date restricted			each other; synthesise these results into overall interpretation of the phenomenon.
20.	Jensen LA, Allen MN. A synthesis of qualitative research on wellness-illness. Qualitative Health Research. 1994; 4: 349-369	Nursing research on health, disease, wellness and illness	Databases: Not specified Keywords: Not specified Other search strategies: Not specified Comprehensiveness: Published between 1980 and 1991	112/n.d.	No restrictions were imposed on the scientific merit of the research, to avoid eliminating germane data	Meta-ethnography
21.	Kearney MH. Enduring love: a grounded formal theory of women's experience of domestic violence. Research in Nursing & Health. 2001; 24:270-82.	Women 's responses to violent relationships	Databases: CINAHL, Medline, Sociofile, Social Work Abstracts, Psyclit, and Dissertation Abstracts. Keywords: Not specified Other search strategies: Manual and reference list searches in the English-language literatures of nursing, medicine, sociology, psychology, anthropology, education, social work, and criminal justice. Comprehensiveness: Purposive: Studies using constant comparative techniques. English only. Date restricted.	13/n.d.	Concepts and theoretical relationships that were not convincingly substantiated were only included if strong support was seen elsewhere in the pooled data. The extent to which the scope of publications was limited was considered. Methodological integrity and completeness of studies were scrutinised.	Grounded formal theory analysis based on Glaser and Strauss 1967 and Strauss and Corbin 1998

22.	Kearney MH. Truthful self-nurturing: a grounded formal theory of women's addiction recovery. Qualitative Health Research. 1998;8:495-512.	Women 's addiction and recovery	Databases: Not specified Keywords: Not specified Other search strategies: Not specified Comprehensiveness: Not clear	10/n.d.	Each study was evaluated using original and more recent criteria for adequacy of grounded theories.	Grounded formal theory
23.	Kennedy HP, Rousseau AL, Low, LK. An exploratory metasynthesis of midwifery practice in the United States. Midwifery 2003; 19: 203-214.	Process and practice of midwifery care	Databases: Not specified Keywords: Not specified Other search strategies: Not specified Comprehensiveness: Authors' own papers	6/n.d.	Not described	Meta-synthesis using meta-ethnography
24.	McNaughton DB. A synthesis of qualitative home visiting research. Public Health Nursing. 2000; 17: 405-414	Home visiting practices of public health nurses	Databases: Medline or CINAHL Keywords: Specified Other search strategies: Reports from reference lists and manual search of selected journals (e.g. Public Health Nursing) Comprehensiveness: Comprehensive. English only. Date restricted	14/17	Not described	Miles and Huberman's principles of qualitative data analysis, including cross case analysis
25.	Meadows-Oliver M. Mothering in public: a meta-	Homeless mothers with	Databases: Not specified	18/n.d.	Not described	Meta-ethnography

	synthesis of homeless women with children living in shelters. Journal for Specialists in Pediatric Nursing. 2003;8:130-6	children living in shelters	Keywords: Not specified Other search strategies: Not specified Comprehensiveness: Not clear. English only.			
26.	Morse JM. Responding to threats to integrity of self. ANS Advances in Nursing Science. 1997;19:21-36.	Experiences of patients and their significant others of injury and acute and chronic illness, focusing on threat to integrity of the self	Databases: Not specified Keywords: Not specified Other search strategies: Not specified Comprehensiveness: Studies conducted by the author or her students	9/n.d.	Not described	Qualitative meta-analysis, based on grounded theory approach
27.	Murray J, Ashworth R, Forster A, Young J. Developing a primary care-based stroke service: a review of the qualitative literature. 2003; British Journal of General Practice 53:137-42.	Self-reported experiences of stroke-related, long-term problems	Databases: MEDLINE, AMED, CINAHL, Sociological Abstracts, PsycINFO Keywords: specified Other search strategies: Handsearching, communication with researchers. Comprehensiveness: Comprehensive. English only. Date restricted	23/139	Not described	Main findings were extracted by two reviewers and grouped according to congruency, using a clustering method. Emergent domains were interpreted and labelled, and subdomains identified.
28.	Neill S (2000) Acute childhood illness at home: the parents' perspective. Journal of Advanced Nursing. 2000;	Acute childhood illness at home from the	Databases: CINAHL, MEDLINE, ASSIA and PSYCHLIT (sic).	22/n.d.	Used method of critiquing based on Sandelowski's stages of data analysis and	Method based on content analysis. Key findings and major points of critique

	31: 821-822.	parents' perspective	<p>Keywords: specified</p> <p>Other search strategies: regular hand searching of current issues of journals in local libraries and sourcing of secondary references</p> <p>Comprehensiveness: Comprehensive. English only. No date restrictions</p>		Gould's critiquing strategy.	recorded on index cards. Collation of main findings, visual search for key themes, further development and integration of themes and sub-themes.
29.	Nelson, AM. A metasynthesis: Mothering other-than-normal children. Qualitative Health Research 2002; 12: 515-530.	Mothering other-than-normal children	<p>Databases: CINAHL (sic), Medline, PsychINFO (sic), ERIC, Sociological Abstracts, and Dissertation Abstracts</p> <p>Keywords: Not specified</p> <p>Other search strategies: No details</p> <p>Comprehensiveness: Comprehensive. English only. Date restricted</p>	12/17 [79]	Not described	Meta-ethnography
30.	Nelson, AM. Transition to motherhood JOGNN-Journal of Obstetric Gynecologic and Neonatal Nursing 2003;32: 465-477.	Transition to motherhood	<p>Databases: CINAHL, MEDLINE, PsycINFO, Sociological Abstracts, and Dissertation Abstracts</p> <p>Keywords: Specified</p> <p>Other search strategies: Not specified</p> <p>Comprehensiveness: Limited to studies of maternal transition among women in North America</p>	9/n.d.	Studies were evaluated for inclusion on the basis of their focus and the methodological comparability of their findings. Explicitly describe why the decision was made not to exclude studies on the basis of quality.	Meta-ethnography

			or Australia			
31.	Paterson BL, Thorne S, & Dewis M. Adapting to and managing diabetes. Image: Journal of Nursing Scholarship. 1998;30:57-62.	Adapting to and managing diabetes	<p>Databases: Sociofile, Psychlit, Dissertation Abstracts, CINAHL, MEDLINE, Allied Health) and the Canadian Nurses Association library searched for published and unpublished dissertations and theses</p> <p>Keywords: Not specified</p> <p>Other search strategies: Refereed nursing, social science, and allied health journals and books dated January 1980 through July 1996 searched for articles and chapters.</p> <p>Comprehensiveness: Comprehensive. English only. Date restricted</p>	38/n.d.	Not described	Meta-ethnography
32.	Paterson, BL (2001) The shifting perspectives model of chronic illness. Journal of Nursing Scholarship 2001; 33: 21-26.	Perspectives of chronic illness	<p>Databases: "Computer searches"</p> <p>Keywords: No details</p> <p>Other search strategies: Citation indexes, reference lists, reviews of research, and professional and journal networks.</p> <p>Comprehensiveness: Comprehensive. English only. Date restricted</p>	292/n.d.	Each primary research report was reviewed by at least three members of the research team, using a "standardized appraisal form".	Metastudy, involving: meta data analysis based on meta-ethnography; meta-method; meta-theory
33.	Roberts, KA, Dixon-Woods,	Factors	Databases: Medline, IBSS,	11/n.d.	Study quality not used	Content analysis

	M, Fitzpatrick, R, Abrams, KR & Jones, DR. Factors affecting uptake of childhood immunisation: a Bayesian synthesis of qualitative and quantitative evidence. Lancet 2002; 360:: 1596-1599	affecting uptake of childhood immunisation	Embase, SocSci Abs, PsychLit, Sociofile, and ASSIA databases, Keywords: Not specified Other search strategies: relevant major journals, and article reference lists for studies published in peer-reviewed journals Comprehensiveness: Comprehensive. Date restricted.		as an inclusion criterion	
34.	Russell CK, Bunting SM, Gregory DM. Protective care-receiving: the active role of care-recipients. Journal of Advanced Nursing 1997; 25: 532-540	Role of care recipients within caregiver-care recipient relationships	Databases: Not specified Keywords: Not specified Other search strategies: Used authors' own studies only. Comprehensiveness: Only authors' studies used	3/n.d.	Not described	Meta-ethnography
35.	Sandelowski M, Barroso J. Motherhood in the context of maternal HIV infection Research in Nursing and Health. 2003;26:470-82.	Motherhood in the context of maternal HIV infection	Databases: Not specified Keywords: Not specified Other search strategies: Not specified Comprehensiveness: Comprehensive	56/114	Reports were not excluded for reasons of quality. A posteriori analyses were conducted to determine how reports contributed to the synthesis.	Qualitative metasummary, involving data extraction data abstraction, calculation of frequency effect sizes. Metasynthesis involved taxonomic analysis, constant comparison, reciprocal translation, synthesis of

						in vivo concepts, and use of imported concepts.
36.	Sandelowski M, Lambe C, Barroso J. Stigma in HIV-positive women. Journal of Nursing Scholarship 2004;36:122-8	Stigma and HIV positive women	Databases: Not described Keywords: Not specified Other search strategies: Not described Comprehensiveness: Not clear	93/n.d.	Formal quality appraisal not conducted, but publications were excluded from the review if: violations of rights of human subjects were evident; findings were not supported by any data; reports contained no findings or researcher interpretations.	Metasummary Metasynthesis
37.	Sandelowski, M and Barroso, J. Toward a metasynthesis of qualitative findings on motherhood in HIV-positive women. Research in Nursing and Health 2003; 26: 153-170	Motherhood in the context of HIV infection	Databases: Not specified Keywords: Not specified Other search strategies: Not specified Comprehensiveness: Comprehensive.	45/99	No report was excluded for reasons of quality, because of lack of consensus about quality or the use of quality criteria in systematic reviews. However reports that violated the rights of human subjects were excluded, as were reports that contained "no findings". Selected findings in included reports were excluded if there were no data to support them.	Qualitative metasynthesis, involving transformation of the findings into conceptual form and creation of a taxonomy; use of sustained comparisons; translating invivo concepts using reciprocal translation; and using imported concepts.
38.	Sherwood, G. Meta-synthesis of qualitative	Clients' perspectives on	Databases: Not described	16/n.d.	Standards for scientific rigour based on Burns	Metasynthesis

	analyses of caring: Defining a therapeutic model of nursing. Advanced Practice Nursing Quarterly 1997; 3: 32-42.	caring	Keywords: Not described Other search strategies: Not described Comprehensiveness: Not described		(1989) and Roberts and Burke (1989) were applied to each study. None was eliminated.	
39.	Thorne S, Paterson B, Acorn S, Canam C, Joachim G, Jillings Chronic illness experience: insights from a meta-study. Qualitative Health Research 2002; 12: 437–452.	Chronic disease experience	Databases: CINAHL, PsychLit, Sociofile and MEDLINE Keywords: Not specified Other search strategies: Hand searches of journals in nursing, medicine, psychology, sociology, medical anthropology and rehabilitation sciences; “backtracking from citations”. Comprehensiveness: Comprehensive. English only. Date restricted	292/1000+	Developed an instrument to capture features such as sample characteristics, methodological applications, evidence of inductive analysis, fieldwork strategies, grounding in illustrative data, accountability measures, and auditability.	Metastudy, involving three distinct analytic phases: metatheory, meta-method, meta-data analysis.
40.	Thorne S, Paterson B. Shifting images of chronic illness. Image: Journal of Nursing Scholarship 1998; 30: 173-178	Constructions of the nature of health care relationships in regard to chronic illness	Databases: Sociological Abstracts, Psychological Abstracts, Dissertation Abstracts, CINAHL, MEDLINE, Allied Health Keywords: Not specified Other search strategies: Searches of refereed nursing, social science, and allied health journals and books. Comprehensiveness:	158/400+	Standardised checklist of criteria to evaluate data collection and analysis procedures, research findings, and theoretical frameworks and emergent theory. Evaluation criteria as suggested by Burns (1989) was used to assess the rigour, epistemologic soundness, and	Meta-ethnography

			Comprehensive. English only. Date restricted		fruitfulness of the research methods used. The researchers met to reach consensus regarding their review. Not clear if studies were excluded if they were judged to be low quality.	
41.	Varcoe, C., Rodney, P, McCormick, J. Health care relationships in context: An analysis of three ethnographies. Qualitative Health Research 2003; 13 : 957-973.	Dynamics of health care experience	Databases: Not specified Keywords: Not specified Other search strategies: Authors used their own three studies Comprehensiveness: Authors used their own three studies only	3/n.d.	Not described	Meta-ethnography
42.	Walter FM, Emery J, Braithwaite D, Martheau TM. Lay understanding of familial risk of common chronic diseases: a systematic review and synthesis of qualitative research. Annals of Family Medicine 2004; 2: 583-594	Understandings of familial risk held by people with a family history of cancer, coronary artery disease, and diabetes mellitus.	Databases: Medline, Web of Science, Psycinfo, CancerLit, Embase, Cinahl, Sigle, and Sociofile. Keywords: Specified Other search strategies: Reference chaining, contact with authors. Comprehensiveness: Comprehensive	11/240	Modified version of CASP "appraisal scoring system". It is not clear whether any papers were excluded after the quality appraisal or how the appraisal affected the synthesis.	Meta-ethnography

Note: grey literature refers to material that is not formally published (e.g. internal reports).

References

- Bravata, D.M., McDonald, K.M., Shojania, K.G., Sundaram, V., Owens, D.K. (2005). 'Challenges in systematic reviews: synthesis of topics related to the delivery, organization, and financing of health care.' *Annals of Internal Medicine* 142 (12 part 2): 1056-65
- Booth, A., O'Rourke, A.J. (1997) 'The value of structured abstracts in information retrieval from MEDLINE.' *Health Libraries Review* 14: 157-166
- Booth A (In press) "'Brimful of STARLITE": Towards standards for literature searches. Journal of the Medical Library Association'
- Chapple, A, Rogers, A (1999) 'Self-care and its relevance to developing demand management strategies: a review of qualitative research', *Health and Social Care in the Community*, 6: 445-454.
- Cook, DJ, Meade, MO & Perry, AG (2001) 'Qualitative studies on the patient's experience of weaning from mechanical ventilation', *Chest* 120, 6: 469S-473S.
- Chalmers, I., Hedges, HV., Cooper, H. (2002) 'A brief history of research synthesis.' *Evaluation and the Health Professions* 25: 12-37
- Christensen, E. (2001) 'Quality of reporting of meta-analyses: the QUOROM statement. Will it help?' *Journal of Hepatology* 34: 342-345.
- Dixon-Woods M, Sutton AJ, Shaw RL, Miller T, Smith J, Young B, Bonas S, Booth A, Jones DR. (in press) 'Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods.' *Journal of Health Services Research and Policy*
- Dixon-Woods, M, Agarwal, S, Jones, D, Young, B, Sutton, A (2005), 'Synthesising qualitative and quantitative evidence: a review of possible methods' *Journal of Health Services Research and Policy* 10: 45-53.
- Dixon-Woods, M., Bonas, S., Booth, A., Jones, D., Miller, T., Shaw, R., Smith, J.A., Sutton, A., Young, B. (2006a) 'How can systematic reviews incorporate qualitative research? A critical perspective', *Qualitative Research* 6: 27-44
- Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, Hsu R, Katbamna S, Olsen R, Smith LK, Riley R, Sutton AJ (2006b) 'Conducting a critical interpretive review of the literature on access to healthcare by vulnerable groups.' *BMC Medical Research Methodology* 6: 35
- Dixon-Woods, M., Fitzpatrick, R., Roberts, K. (2001), 'Including qualitative research in systematic reviews: opportunities and problems', *Journal of Evaluation in Clinical Practice* 7: 125-33.
- Dixon-Woods, M, Shaw, R.L., Agarwal, S., Smith, J.A. (2004) 'The problem of appraising qualitative research' *Quality and Safety in Healthcare* 13:223-225.
- Egger M, Davey Smith G, O'Rourke K. (1995) 'Rationale, potentials and promise of systematic reviews.' In: Chalmers I, Altman DG (eds) *Systematic reviews* London: BMJ Publishing Group
- Elliott, J. (2005) 'Using narrative in social research: qualitative and quantitative approaches.' London: Sage

- Greenhalgh, T., Peacock, R. (2005) 'Effectiveness and efficiency of search methods in systematic review of complex evidence: audit of primary sources.' *British Medical Journal* 331: 1064-1065
- Greenhalgh, T., Robert, G., Macfarlane, .F, Bate, P., Kyriakidou, O., Peacock, R. (2005) 'Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review.' *Social Science and Medicine* 61, 417-430
- Hemels, ME., Vicente, C., Sadri, H., Masson, M.J., Einarson, T.R. (2004), 'Quality assessment of meta-analyses of RCTs of pharmacotherapy in major depressive disorder' *Current Medical Research and Opinion* 20: 477-484.
- Klassen, TP., Jadad, AR., Moher, D. (1998) 'Guides for reading and interpreting systematic reviews: I: getting started'. *Archives of Pediatric and Adolescent Medicine* 152: 622-623
- McAlister, FA., Clark, HD., van Walraven, C., Straus SE., Lawson, FM., Moher, D., Mulrow CD. (1999) 'The medical review article revisited: has the science improved?' *Annals of Internal Medicine*.131: 947-51
- Pidgeon, N., Henwood, K. (2004). 'Grounded theory'. In: Hardy M, Bryman A (eds) *Handbook of Data Analysis* London: Sage
- Mays, N., Pope, C., Popay, J. (2005), 'Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field', *J Health Serv Res Policy*. 10 Suppl, 1, 6-20.
- Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D. for the QUORUM group. (1999), 'Improving the quality of reporting of meta-analysis of randomised controlled trials: the QUORUM statement.' *The Lancet* 354: 1896-1900.
- Mulrow, C. (1987)' The medical review article: state of the science.' *Annals of Internal Medicine* 106: 485-488
- Noblit, G.W., Hare, R.D. (1988), *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park CA, Sage.
- Paterson, B., Thorne, S., Canam, C., Jillings, C. (2001), '*Meta-study of qualitative research: A practical guide to meta-analysis and meta-synthesis*', Thousand Oaks CA, Sage.
- Peters, J., Sutton, A.J., Jones, D.R., Rushton, L., Abrams, K.A. (in press) 'Systematic reviews and meta-analyses of animal toxicology studies: a systematic review of the literature.' (Accepted: *Journal of Environmental Science and Health*)
- Schutz, A (1962) '*Collected papers, Vol. 1*', The Hague, Nijhoff.
- Shaw, RL, Booth, A, Sutton, AJ, Miller, T, Smith, JA, Young, B, Jones, DR, Dixon-Woods, M (2004), 'Finding qualitative research: an evaluation of search strategies', *BMC Medical Research Methodology*.16, 4-5.
- Shea, B, Dube, C, Moher, D. (2000) 'Assessing the quality of reports of systematic review: The QUORUM statement compared to other tools.' In: Egger, M, Davey Smith, G, Altman, DG, editors. *Systematic reviews*. 2nd ed. BMJ Publishing Group: London.
- Sheldon, T.A. (2005), 'Making evidence synthesis more useful for management and policy-making', *Journal of Health Services Research and Policy*,10 Suppl 1:1-5.
- Smith, JT. (1996) 'Meta-analysis: the librarian as a member of an interdisciplinary research team.' *Library Trends* 45: 265-79

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., et al. (2000), 'Meta-analysis of observational studies in epidemiology: a proposal for reporting', *JAMA* 283: 2008-2012

Thorne, S., Jensen, L., Kearney, M.H., Noblit, G., Sandelowski, M. (2004), 'Qualitative metasynthesis: reflections on methodological orientation and ideological agenda', *Qualitative Health Research*, 14:1342-1365

Appendix D - Paper M3

**This is the Post-print version with changes from referees' comments.
Embargoed until October 2013. Publisher copyright and the source are
acknowledged as follows:**

**Carroll C, Booth A, Lloyd-Jones M. Should we Exclude Inadequately-reported
Studies from Qualitative Systematic Reviews? An Evaluation of Sensitivity
Analyses in Two Case Study Reviews, *Qualitative Health Research* 2012; 22 (10):
1425-1434.**

Should we Exclude Inadequately Reported Studies From Qualitative Systematic Reviews? An Evaluation of Sensitivity Analyses in Two Case Study Reviews

Christopher Carroll¹, Andrew Booth¹ and Myfanwy Lloyd-Jones¹

¹University of Sheffield, Sheffield, UK

Corresponding author:

Christopher Carroll, University of Sheffield, Regent Court, Regent Street, Sheffield, S1
4DA, UK

Email: c.carroll@sheffield.ac.uk

Abstract

The role of critical appraisal of qualitative studies in systematic reviews remains an ongoing cause for debate. Key to such a debate is whether quality assessment can or should be used to exclude studies. In our study, we extended the use of existing criteria to assess the quality of reporting of studies included in two qualitative systematic reviews. We then excluded studies deemed to be inadequately reported from the subsequent analysis. We tested the impact of these exclusions on the overall findings of the synthesis and its depth or thickness. Exclusion of so-called inadequately reported studies had no meaningful effect on the synthesis. There was a correlation between quality of reporting in a study and its values as a source for the final synthesis. We propose that there is a possible case for excluding inadequately reported studies from qualitative evidence synthesis.

Keywords

critical methods; qualitative analysis; research, mixed methods; systematic reviews; validity

The internal validity of a systematic review is dependent on both the quality of included studies and the reliability of their findings. The exact meaning of both quality and reliability in the context of quality assessment is vigorously contested. Debate is especially vocal when the review evidence is qualitative (Barbour, 2001; Eakin & Mykhalovskiy, 2003; Popay, Rogers, & Williams, 1998). For example, the heavy reliance on direct observation in anthropology and ethnography has led some commentators to caution against any attempt to assess the quality of such research indirectly using predetermined criterion based checklists (Power, 2001).

Notwithstanding such reservations many of those involved in research synthesis, ourselves included, take a pragmatic and fundamentally utilitarian stance toward the potential contribution of qualitative research. Indeed calls for just such a pragmatic approach have recently issued from this very journal (Thorne, 2011). We contend that, if findings from individual qualitative data studies are to contribute to a collective understanding of a particular phenomenon, then the resulting synthesis must be based on how the original researchers report their findings. Furthermore, even though reviewers might prefer to access insights from the wider context within which the research has been conducted, in the absence

of such insights they can only base assessments of internal coherence and technical consistency on the published accounts of such research (Sandelowski & Barroso, 2007).

Currently, there is much debate and little consensus around the feasibility and usefulness of the quality assessment of qualitative studies in evidence synthesis (Dixon-Woods, Bonas, et al., 2006; Dixon-Woods, Shaw, Agarwal, & Smith, 2004; Dixon-Woods, Sutton, et al., 2007; Hannes, Lockwood, & Pearson, 2010; Lincoln, 1995; Mays & Pope, 1995). In some techniques, such as meta-ethnography (Campbell et al., 2003), critical interpretive synthesis (Barnett-Page & Thomas, 2009) and framework synthesis (Brunton, Oliver, Oliver, & Loren, 2006; Oliver et al., 2008), studies might be excluded explicitly on the basis of the quality assessment. In other examples of qualitative evidence synthesis quality assessment has not been used at all (Gomersall, Madill, & Summers, 2011).

Consequently, researchers have called for both empirical research and theoretical debate to address important questions about the purpose of quality assessment in such types of evidence synthesis (Dixon-Woods, Sutton, et al., 2007). Key to the controversy surrounding quality assessment is an understanding of the effects of including articles of differing quality within an interpretive synthesis. It is generally agreed that some form of quality assessment is required to identify flawed research that might distort a review's findings (Dixon-Woods & Fitzpatrick, 2001; Dixon-Woods, Shaw, et al., 2004). We therefore aimed to assess whether excluding those studies that inadequately report their methods demonstrates any effect on the findings of qualitative evidence synthesis. By examining the effect of exclusion of studies on the basis of the adequacy of reporting of methods we sought to advance the ongoing debate concerning quality assessment in qualitative evidence synthesis.

Methods

The Data

We had previously performed two systematic reviews of qualitative data on the following topics: young people's views relating to school sexual health services (Carroll, Lloyd-Jones, Cooke, & Owen, 2012); and health professionals' views and experiences of online education techniques (Carroll, Booth, Papaioannou, Sutton & Wong, 2009). Both systematic reviews included so-called "views" studies. According to Harden et al. (2004, p.794), views studies are studies that, "attempt to understand . . . issues from the perspectives of the people they affect." Authors of such studies place "people's own voices at the centre of their analysis". Previous published reviews of people's views have included studies that use a wide variety of methods. Data collection methods used by such studies included interviews and focus groups alongside questionnaires, which use frequencies to quantify the proportion of people with a particular view or preference (Harden et al., 2004). Our own systematic reviews included studies in which authors had employed a similar variety of methods. Data could be structured or unstructured, were often textual, and described people's own, personal, subjective experiences or views of the service or intervention of interest.

We extracted these data from the Results sections of included studies and the data were analyzed qualitatively (Dixon-Woods, Agarwal, Jones, Young, & Sutton, 2005). We chose a grounded theory, inductive approach to data analysis in both reviews; namely, secondary thematic analysis (Miles & Huberman, 1984). We used this reductive approach in classifying the extracted data into themes. Themes that related to each other were placed under a new, broader theme. The resulting thematic framework reflected the experiences and views of participants toward the phenomena of interest and was based on our own interpretation of the data. Our synthesis involved interpreting and integrating, rather than aggregating, findings from multiple studies. We excluded no study from either review on the basis of quality.

The Quality Assessment Process

We derived a simple checklist for quality assessment based on four questions relating to key procedural elements of research. These criteria have previously appeared as elements of other qualitative research checklists, tools and discussion papers (Dixon-Woods, Shaw, et al., 2004; Mays & Pope, 1995). For example these four questions represent items three, four, five and seven from the Critical Appraisal Skills Programme (CASP) tool (Public Health Research Unit, 2006) and items 1, 15, 25, 30 and 31 from the Evaluation Tool for Qualitative Studies (ETQS; Health Care Practice Research & Development Unit, 2009). However in making our own assessment we focused only on how adequately each methodological issue was addressed by the descriptions presented in each included article (see Table 1). In other words, we assessed only the text describing these elements, rather than attempting to appraise the actual conduct of each study, which is more typically the intention behind appraising qualitative research (Dixon-Woods, Shaw, et al., 2004; Hannes et al., 2010; Mays & Pope, 1995; Whittemore, Chase, & Mandle, 2001). We believe that our empirical study represents the first practical attempt to evaluate the value of each study to a systematic review by explicitly and solely assessing the adequacy with which procedural elements are described in a study.

INSERT TABLE 1 ABOUT HERE

We took the decision to focus on quality of reporting for two reasons. First, researchers have pointed out previously that any appraisal checklist essentially only assesses what is reported in a publication (Dixon-Woods et al., 2004). The limitations of judging quality on the basis of a published account applies equally to all types of research. Debates on criteria acknowledge all too infrequently that we cannot really begin to assess anything about a study unless it is adequately reported (Dixon-Woods, Shaw, et al., 2004; Mays & Pope, 1995). Arguably it is not possible to assess the validity of a study, for example, in terms of credibility, truthfulness, authenticity, believability and so forth, if authors do not report, or

only report inadequately, the information required to make such a judgment. Such information might include the authors' theoretical perspectives, how and why the data were collected, the application and appropriateness of any validation tests, and the relationship between the authors' interpretations and their data.

Despite opinions to the contrary (Hannes et al., 2010), the quality of reporting is a determinant of any assessment of methodological soundness. Reviewers can only apply assessment criteria once they have established that the analysis and findings have been reported transparently (Mays & Pope, 1995). We therefore chose to focus on the auditability and transparency of the methods of each study, as reported in the publication, because this intuitively seemed a good place to start. We did not evaluate whether the methods described were either appropriate or well conducted. We only assessed whether the methods were reported in adequate detail. We acknowledge the possibility that inadequately reported studies can be well conducted and can be used to offer important insights (Dixon-Woods, Shaw, et al., 2004; Hannes et al., 2010). Nevertheless, a reviewer must be equipped with adequate information to make such an assessment. Even though reporting of primary research studies may be constrained by limited word counts and restrictions imposed by journal formats these should not be allowed to mitigate in favor of their potential quality. A reviewer cannot afford to be forgiving when authors fail to report how they chose the study design or selected participants, or how they collected and analyzed their data.

The second reason for adopting the chosen approach was recognition that elements of a study relating to reporting of methods are more easily judged and apprehended than other study features. Assessment of such elements consists simply of determining whether each publication clearly describes the question and study design; how the participants were recruited or selected; and the methods of data collection and analysis used (Table 1). Similar criteria have been used by other systematic reviewers in more extensive quality assessments

of views studies (Barnett-Page & Thomas, 2009; Thomas & Harden, 2008). The same criteria have also figured in lists of prompts and other checklists for the consideration of qualitative research (Dixon-Woods, Shaw, et al., 2004; Mays & Pope, 1995). We extended their use here by utilizing them as exclusive assessment criteria, rather than merely as prompts with which to begin to critique a piece of research.

The relatively small number of criteria described above has universal application to published research. They may in fact be more practical than checklists with greater numbers of questions. More extensive checklists have been found to generate low inter-rater reliability scores even among experienced qualitative systematic reviewers (Dixon-Woods, Sutton, et al., 2007). Two of us independently applied these criteria to all included studies in each review. We assigned definitions to these criteria to make them more easily understood and to minimize the likelihood of subjective judgments by assessors. Our focus on how methods are reported in a publication meant that we did not need to make potentially disputable judgments on such contested issues as researcher bias and validity (Dixon-Woods, Sutton, et al., 2007; Hannes et al., 2010).

Such an approach might seem to treat qualitative research or views studies as a unified body of work, which they clearly are not (Dixon-Woods, Bonas, et al., 2006; Dixon-Woods, Shaw, et al., 2004). Our aim was to be reductive; to simplify for practical purposes. We did not attempt to evaluate the validity or test-retest reliability of this brief checklist, or to compare its technical performance to that of existing published checklists. We considered this unnecessary given that this checklist embodies questions already commonly used in many existing checklists. We simply chose to focus explicitly and exclusively on the single domain of reporting or description of each study.

In conducting an assessment, the primary reviewer (Christopher Carroll) read the original publication and extracted any text that addressed the quality assessment questions,

where available, into the checklist form. Such text was principally identified from the Introduction and Methods sections, or their equivalents, of each publication. The reviewer then assigned an answer of Yes, No, or, in cases of uncertainty, Unclear against each criterion. A second reviewer (Andrew Booth or Myfanwy Lloyd-Jones) then validated or challenged the assessment by examining both the extracted text and the original publication before arriving at their own judgement.

We then dichotomized studies into “adequately reported” or “inadequately reported” groups. The review team decided that studies that had been assigned a clear Yes against two or more criteria (i.e., the publication clearly satisfied at least two of the key quality criteria) would be categorized as adequately reported studies. Conversely, where a study was assigned only a single clear Yes response (i.e. where only one of study design, recruitment, data collection and analysis was adequate), or where it received no Yes responses at all, it was categorized as inadequately reported. For examples of adequately reported and inadequately reported categorizations, see Figure 1.

INSERT FIGURE 1 ABOUT HERE

The reviewers discussed any differences of opinion when assigning criteria to studies and reached a consensus on categorization to one of the two groups. We dichotomized studies into adequately and inadequately reported studies because the scale used (either zero or one, or two, three or four) accommodated a binary outcome of include or exclude for each study and simplified the subsequent sensitivity analysis.

Sensitivity Analysis

We then performed a sensitivity analysis for each review in which the inadequately reported studies were excluded from the analysis. We assessed whether, and, if so, to what extent, the synthesis was affected by exclusion of these studies (Downe, 2008; Sandelowski, Barroso, & Voils, 2007). First we evaluated whether any of the themes generated in the original

syntheses were lost because of the exclusion of these studies. Then we assessed whether exclusion of studies affected the composite “thickness” of detail (Popay J et al., 1998) or richness of information (Patton, 1990) within the synthesis. In other words, we wished to identify where a theme remained but at the expense of its complexity, richness or dissonance, i.e., the presence of alternative points of view and perspectives (Paterson, Thorne, Canam, & Jillings, 2001).

By examining actual findings, and the degree to which they contributed to the final synthesis, we did not privilege how methods in included studies were reported over their findings. Instead we took both methods and findings into account. We conducted these sensitivity analyses to examine whether we had introduced a possible bias in favour of the procedural elements of the constituent research. In examining for such a bias we hoped to counter the oft cited criticism of quality assessment methods, namely that studies of low methodological quality can nevertheless be the source for novel insights not provided by adequately reported studies (Dixon-Woods, Bonas, et al., 2006; Pawson, 2006).

Results

Only 10 of the 19 included studies in the review of young people’s attitudes toward school sexual health services were adequately reported. Nine studies were judged to be inadequately reported and these were excluded from the synthesis for the purposes of our sensitivity analysis. The exclusion of such a large number of studies had a negligible impact. We used thematic synthesis to generate eight principal themes reflecting factors affecting young people’s use or non-use of the services in question (Carroll et al., 2012). Each of these principal themes emerged from other themes generated from the primary studies. No single principal theme was completely dependent on data from inadequately reported studies. For example, although 13 of the 19 studies were the source for the theme of confidentiality and

disclosure only three of these were inadequately reported studies. No additional data emerged as exclusive findings from the inadequately reported studies.

All of the themes identified in this review of sexual health studies contained comparable ratios of contribution from the adequately and inadequately reported studies. None of the themes were affected by exclusion of studies for the sensitivity analysis. Five of the nine excluded studies (Emihovich & Herrington, 1997; Guttmacher et al., 1995; Kirby et al., 1999; Nelson & Quinney, 1997; Zabin, Stark, & Emerson, 1991) were the source for only to one or two of the eight themes, with two being the source for only three of the eight themes (Schuster, Bell, Berry, & Kanouse, 1997; Zeanah et al., 1996). The two remaining studies were the source for data for five and six themes respectively (Tanner, Kirton, Stone, & Ingham, 2003; Washkansky, 2008). Findings from these ubiquitous studies contributed little in terms of variety, dissonance, or a novel perspective within each of the themes. The limited contribution derived from these two studies reflected the fact that their data largely mirrored that reported by adequately reported studies, and thus we interpreted them in the same way to generate the same themes.

Limited additional richness was provided by data derived from the less adequately reported studies. For example, participants in one excluded study made the point, not expressed elsewhere, that the gender of staff was important to service users (Guttmacher et al., 1995). In another study young people expressed a preference for provision of comprehensive health services compared with sexual health services alone, not because of concerns about accessing the latter, but because they simply wanted easy access to more comprehensive healthcare (Zeanah et al., 1996).

Data from adequately reported studies, by contrast, was markedly more substantial and richer. For example, seven of these 10 studies were the source for between five and all eight of the themes. What is more noteworthy, however, is that we found that, with the

exception of the two cases cited above, all instances of dissonance, richness, or complexity for each theme emerged from one or more of these adequately reported studies. For example, the barrier to service use presented by personal anxiety about disclosure, and the facilitator represented by users' confidence in the levels of privacy was described in five adequately reported studies. Thus each of these studies proved a source for a more rounded out or balanced perspective on the same phenomenon.

The review of experiences of online learning among United Kingdom health professionals also demonstrated a negligible effect of exclusions. We excluded nine of 19 studies from the analysis on the basis of inadequate reporting of methods used (Carroll et al., 2009). None of the 10 subthemes or five principal themes generated from the data using secondary thematic analysis depended exclusively on the inadequately reported studies. We found that one excluded study was a source for every principal theme (Anthony & Duffy, 2003) and another study for four of the five themes (Kinghorn, 2005). Of the remaining seven inadequately reported studies, two were the source for a single theme (Hare, Davis, & Shepherd, 2006; Hurst, 2005).

By comparison three of the 10 adequately reported studies contributed to every theme (Hall, Harvey, Meerabeau, & Muggleston, 2004; Whittington, Cook, Barratt, & Jenkins, 2004; Wilkinson, Forbes, Bloomfield, & Fincham, 2004;). A further three studies were the source for three out of the five themes (Conole, Hall, & Smith, 2002; Gresty, Skirton, & Evenden, 2007; Larsen & Jenkins, 2005). Only one adequately reported study was the source for only a single theme (Thorley, Turner, Hussey, Hall, & Agius, 2007). Nevertheless some of the richness of the synthesis was generated from data from the inadequately reported studies in the e-learning review. Seven of the eight studies that were focused exclusively on the online learning experience of nurses were included in the nine inadequately reported studies. Excluding these studies from the synthesis would have resulted in the loss of

valuable data from, and about, nurse learners. As a consequence differences between nurses and other groups, most notably doctors, might have been concealed.

Discussion

Our investigation enabled us to examine several issues. We were able to explore the extent of the contribution from individual studies to a synthesis, based on the adequacy of the description of their basic methods; whether multiple themes were present in individual studies (Sandelowski et al., 2007); and whether the synthesis was adversely affected by excluding so-called inadequately reported studies. We also performed sensitivity analysis for a third review (Carroll, Booth, & Cooper, 2011) but, as no article was excluded, we have not included findings from this review as an additional case study. The main contribution of our sensitivity analyses was to identify the potential omission of findings relating to one particular professional group, found to predominate in one set of inadequately reported studies. With the exception of this observation, our sensitivity analyses revealed that exclusion of inadequately reported studies from the syntheses did not affect the findings in any meaningful way. That is to say, no theme or subtheme generated by either of the syntheses depended on those studies with the most limited reporting of methodology.

Further examination of the contribution made by inadequately reported studies indicated that they tended to lack thickness of detail in comparison to the adequately reported studies, and thus contributed little in the way of richness. Simply put, data derived from inadequately reported studies did little to supplement data from adequately reported studies. Such a conclusion remained true whether we judged their contribution in terms of individual constituents to a theme or in terms of different perspectives within the themes or the resultant synthetic model. Such an observation is perhaps not surprising because, if design or methods of participant selection, data collection, or analysis, are not clearly described in a published

study, then that study is unlikely to be a source for findings of more than limited value.

Conversely, themes present in the final syntheses were determined only by data from adequately reported studies.

We did not assess contribution only in terms of the number of studies contributing to the themes present within the resultant model or framework (i.e., thus reducing qualitative data synthesis to a quantitative sensitivity analysis). We also evaluated the richness or thickness of the detail within each theme, including identifying the presence of alternative viewpoints or dissonance. We found that very few inadequately reported studies proved to be the source for novel or diverse contributions that were retained in the subsequent elaboration of the themes. We acknowledge that it is difficult to gauge the exact additional value of the limited number of original insights over those already derived from data in the adequately reported studies. Nevertheless, we did conclude that none of the unique contributions was sufficiently substantial to generate a new theme. If anything, such insights were only able to add nuances to themes that had already emerged from groups of adequately reported studies.

Two review teams have previously reported a lack of specific impact from relatively lower quality studies following sensitivity analyses for their qualitative reviews (Noyes & Popay, 2007; Thomas & Harden, 2008). In these reviews the authors attempted to assess both what was said to be done as well as what was actually done (validity). Both teams reported that the contribution from apparently poorer studies was both less in terms of both material and the depth of the synthesis. We augment these findings by offering an analysis of an additional two reviews. In contrast to these previous analyses which used extensive quality assessment checklists, we focused explicitly and exclusively on criteria associated with the quality of reporting. We describe an approach that represents a relatively straightforward and pragmatic alternative to the lengthy checklists employed for assessment of studies of qualitative data. The simple assessment criteria applied here, requiring only extraction and

evaluation of what is actually described or reported, may also afford a more consistent means of appraisal. Program evaluation research has demonstrated that simple, clearly defined approaches achieve more consistent results than more lengthy, complex, or vague programs or tools with their greater scope for variation (Grol et al., 1998). The process of critical appraisal, especially of qualitative studies, is reported to suffer from similar tendencies (Dixon-Woods, Bonas, et al., 2006; Dixon-Woods, Sutton, et al., 2007). When compared to other appraisal checklists, the assessment system reported and applied here might minimize the potential for appraiser bias (Mays & Pope, 1995). Assessors are not required to make subjective judgments concerning, for example, theoretical perspectives, the link between theory and methods, or the validity, i.e., the authenticity or credibility of findings from a study. Instead they simply identify, extract and assess the actual text relating to the stated criteria.

The research also underlines the practical realities of having to deal with the inadequacies of poor reporting or thin description of qualitative studies. Half of the studies in our reviews were inadequately reported. Such a ratio is comparable to that for other reviews: Noyes and Popay (2007) judged seven out of their 27 included studies to be “thin” and Harden et al. (2004) reported that only four out of their 35 included studies satisfied all seven of their quality criteria. In both these reviews, the reviewers extended quality assessment criteria beyond simple methodological descriptions by also attempting to assess validity. Notwithstanding vigorous academic debate over criteria for, and approaches to, quality assessment (Dixon-Woods, Shaw, et al., 2004; Hannes et al., 2010; Mays & Pope, 1995; Whitemore et al., 2001), the practical reality for the systematic reviewer of qualitative or views studies is that the reporting found in many studies will be inadequate to permit a robust assessment of validity. Consequently the resultant assessment is more often of the reporting than of the validity.

Calls for a Consolidated Standards of Reporting Trials (CONSORT) type (Schulz, Altman, & Moher, 2010) approach to the reporting of qualitative research might serve to address such a situation in the future. Current syntheses will continue to be constrained by the potential inclusion of inadequately reported qualitative or views studies published in the last decade and before. Isolated examples do exist where no studies would be excluded on the basis of quality of reporting. Indeed this was the case for the authors' own experience of a review of people's views of various chemopreventive agents (Carroll et al., 2011) in which all 20 studies met the criteria outlined above. However for inclusion studies had to have been published in 2003 or later. In a similar vein, Noyes and Popay (2007) reported that more recent studies identified by their updated review were better, that is, had greater thickness of detail, than studies identified for the original review. These more recent, thicker studies were also the source for two new themes in the synthesis. Such a finding suggests a possible improvement in reporting or description of qualitative research over the last five years. Limiting retrieval to recent studies might therefore serve as a surrogate quality threshold.

Implications

On the basis of this exploratory research we believe that there is an increasingly strong argument for excluding inadequately reported studies from qualitative systematic reviews. This argument is confirmed by sensitivity analyses across four different reviews, two reported here and two published by other review teams. Several published methods for the synthesis of qualitative studies, such as meta-ethnography, critical interpretive synthesis, and framework synthesis, already advocate exclusion of apparently low quality studies (Barnett-Page & Thomas, 2009). Our findings suggest that a similar approach might also be appropriate when performing secondary thematic analysis or other interpretive (i.e., non aggregative) approaches. Alternatively, where a qualitative systematic review team does not feel able to exclude studies pre-synthesis based on the adequacy of reporting, they should at

least assess the adequacy of the published description of the methods to inform a subsequent sensitivity analysis.

A default position, until these emergent findings are confirmed beyond reasonable doubt, would be to require that reviewers conduct a post synthesis sensitivity analysis to assess whether anything, no matter how apparently insignificant, might have been lost to the synthesis by excluding inadequately reported studies. Such an analysis would gauge the impact of excluding inadequately reported studies on the final synthesis. It would also inform reflections on the robustness of the resulting synthesis. For example, it would allow identification, and subsequent investigation, of instances where a particular finding or group of findings is dependent, either exclusively or disproportionately, on one or more inadequately reported studies. The review team would then be in a position to make an informed and appropriate decision on how they are to handle this. In our e-learning case study, we identified a group of inadequately reported studies that reported the perspectives of a single professional group. Excluding such studies on the basis of reporting quality might have impacted on the external validity of the review findings. Where reviewers feel that external validity has been compromised in such a way they could make an explicit decision whether to retain inadequately reported studies. Sensitivity analysis would also identify where findings from an adequately reported study contradict those from less completely reported alternatives. Techniques for sensitivity analysis remain incompletely specified within qualitative evidence synthesis and so offer a promising target for future empirical work and methodological guidance.

Alternatively, an evaluation of the basic reporting of methods prior to synthesis could become a critical preliminary exclusion stage for every qualitative review. Findings from a synthesis that are clearly supported by methodologically transparent and well-described primary research studies are potentially more robust than those of a synthesis based in part or

in whole on studies for which a validity assessment proves elusive because of an absence of relevant information. Given how problematic it is to evaluate validity in the absence of transparency of reporting, it is likely that thorough assessment is only possible for adequately reported studies anyway.

Limitations

Our research has several limitations. The review team might have encountered different findings had they employed either different synthesis techniques or different quality assessment approaches, or both. Qualitative synthesis is inherently interpretive, so different reviewers might generate slightly different synthetic models from the same data, with differential impact from excluding the inadequately reported studies. There are issues around the reproducibility and validity of the appraisals, as for any such assessment of qualitative studies. We aimed to control for such variability by keeping the criteria simple and defined, and by putting in place procedures to validate independently the judgments made by the primary reviewer. Such validation was followed by discussion and consensus on how studies had been categorized.

It is also possible that, because the sensitivity analyses performed were post hoc, small novel contributions to the syntheses derived from inadequately reported studies were simply absorbed by the power of the interpreted themes, given the potential tendency to seek commonalities rather than dissonance (Petticrew & Roberts, 2006). We initially hoped to base assessments of quality solely on text from the Introduction and Methods sections of the published studies. However it became apparent during the assessments that additional relevant information on methods appeared elsewhere. Some information was contained in authors' own reports of the limitations of their study, typically in the Discussion sections. In such instances, the additional data simply confirmed the categorization based on the earlier data. Nevertheless, we do recommend that future reviewers make a specific attempt to

harvest data from Discussion sections when conducting their preliminary assessment of reporting quality.

Finally, we only focused on the adequacy of descriptions of methods within publications. Reporting of methods is clearly not a proxy for the methodological soundness of a study (Hannes et al., 2010) and a potentially unsound study can receive an adequate assessment following application of our criteria. Nevertheless, we consider such transparency to be a first step towards being able to assess the more fundamental and essential elements that determine the quality of qualitative research. If the reporting of a study is inadequate in the first place, it will prove difficult to apply validity criteria at all.

Conclusion

We extended and applied simple, pragmatic, quality assessment criteria to reports of studies included in two systematic reviews of people's views in topics from public health and the education of health professionals. Our quality assessment focused explicitly on the reporting or description of a small number of clearly-defined elements of research procedure within these qualitative data studies. We then performed a sensitivity analysis to evaluate the impact of excluding the inadequately reported studies from the two syntheses. We found that in no case did these exclusions appear to affect either the overall conceptual findings of these systematic reviews or the richness of the data underpinning their results.

The implications of the above are twofold. Reviewers could apply the given critical appraisal criteria either (pre-synthesis) to exclude inadequately reported studies. Alternatively they could test the robustness of review findings (post synthesis) through sensitivity analyses. Different reviewers working on different topics need to utilize both strategies to assess the value of each to their particular qualitative systematic review. It would also be useful to compare the criteria and approaches described here with other approaches to quality assessment such as that proposed by Sandelowski and Barroso (2007). We submit these

findings as a contribution to the ongoing debate on the critical appraisal and quality assessment of studies within the field of qualitative evidence synthesis.

Acknowledgments

We would like to thank Diana Papaioannou, Anthea Sutton, Jo Cooke, and Katy Cooper.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

References

- Anthony, D., & Duffy, K. (2003). An evaluation of a tissue viability online course. *Information Technology in Nursing (ITIN)*, 15, 20-28.
- Barbour, R. (2001). Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *British Medical Journal*, 322, 1115-1117. doi: 10.1136/bmj.322.7294.1115
- Barnett-Page, E., & Thomas, J. (2009). Methods for the synthesis of qualitative research: a critical review. *BMC Medical Research Methodology*, 9, 59. doi:10.1186/1471-2288-9-59
- Brunton, G., Oliver, S., Oliver, K., & Lorenc, T. (2006). *A synthesis of research addressing children's, young people's and parents' views of walking and cycling for transport for London*. Retrieved from <http://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/WalkingAndCyclingWEB.pdf>
- Campbell, R., Pound, P., Pope, C., Britten, N., Pill, R., Morgan, N., & Donovan, J. (2003). Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social Science & Medicine*, 65, 671-684. doi:10.1016/S0277-9536(02)00064-3
- Carroll, C., Booth, A., & Cooper, K. (2011). A worked example of best-fit framework synthesis: A systematic review of views concerning the taking of potential chemopreventive agents, *BMC Medical Research Methodology*, 11: 29. doi: 10.1186/1471-2288-11-29
- Carroll, C., Booth, A., Papaioannou, D., Sutton, A., & Wong, R. (2009) UK Healthcare professionals' experience of e-learning techniques: a systematic review of qualitative data. *Journal of Continuing Education in the Health Professions*, 29 (4): 235-241. Retrieved from <http://www.jcehp.com/>

- Carroll, C., Lloyd-Jones, M., Cooke, J., & Owen, J. (2012). Reasons for the use and non-use of school sexual health services: A systematic review of young people's views. *Journal of Public Health* (in press). doi:10.1093/pubmed/fdr103
- Conole, G., Hall, M., & Smith, S. (2002). An evaluation of an online course for medical practitioners. *Educational Technology & Society*, 5, 66-75.
- Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research and Policy*, 10, 45-53B. doi: 10.1258/1355819052801804
- Dixon-Woods, M., Bonas, S., Booth, A., Jones, D., Miller, T., Shaw, R., Smith, L., . . . Young, B. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, 6, 27-44. doi: 10.1177/1468794106058867
- Dixon-Woods, M., & Fitzpatrick, R. (2001). Qualitative research in systematic reviews has established a place for itself. *British Medical Journal*, 323, 765-766. doi:10.1136/bmj.323.7316.765
- Dixon-Woods, M., Shaw, R., Agarwal, S., & Smith, J. (2004). The problem of appraising qualitative research. *Quality and Safety in Health Care*, 13, 223-225. doi: 10.1136/qshc.2003.008714
- Dixon-Woods, M., Sutton, A., Shaw, R., Miller, T., Smith, J., Young, B., Bonas, S., . . . Jones, D. (2007). Appraising qualitative research for inclusion in systematic reviews: A quantitative and qualitative comparison of three methods. *Journal of Health Services Research and Policy*, 12, 42-47. doi: 10.1258/135581907779497486
- Downe, S. (2008). Metasynthesis: a guide to knitting smoke. *Evidence Based Midwifery*, 6, 4-8. Retrieved from <http://www.rcm.org.uk/ebm/>
- Eakin, J., & Mykhalovskiy, E. (2003). Reframing the evaluation of qualitative health research: reflections on a review of appraisal guidelines in the health sciences.

Journal of Evaluation in Clinical Practice, 9, 187-194. doi: 10.1046/j.1365-2753.2003.00392.x

Emihovich, C., & Herrington, C. (1997). *Sex, kids, and politics. Health services in schools*. New York: Teachers College Press.

Gomersall, T., Madill, A., & Summers, L. (2011). Metasynthesis of the self-management of Type 2 diabetes. *Qualitative Health Research*, 21, 853-871. doi:10.1177/1049732311402096

Gresty, K., Skirton, H., & Evenden, A. (2007). Addressing the issue of e-learning and online genetics for health professionals. *Nursing & Health Sciences*, 9, 14-22. doi: 10.1111/j.1442-2018.2007.00296.x

Grol, R., Dalhuijsen, J., Thomas, S., Veld, C., Rutten, G., & Mookink, H. (1998). Attributes of clinical guidelines that influence use of guidelines in general practice: An observational study. *British Medical Journal*, 317, 858-861. doi: 10.1136/bmj.317.7162.858

Guttmacher, S., Lieberman, L., Hoi-Chang, W., Radosh, A., Rafferty, Y. Ward, D., & Freudenberg, N. (1995). Gender differences in attitudes and use of condom availability programs among sexually active students in New York City public high schools. *Journal of the American Medical Womens Association*, 50, 99-102.

Hall, N., Harvey, P., Meerabeau, L., & Muggleston, D. (2004). *An evaluation of online training in the NHS workplace*. Retrieved from http://www.recordingachievement.org/images/pdfs/case_studies/HE5Psection/an%20evaluation%20of%20online%20training%20in%20the%20nhs%20workplace.pdf

Hannes, K., Lockwood, C., & Pearson, A. (2010). A comparative analysis of three online appraisal instruments' ability to assess validity in qualitative research. *Qualitative Health Research*, 2, 1736-1743. doi: 10.1177/1049732310378656

- Hannes, K. (2011). Chapter 4: Critical appraisal of qualitative research. In: J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, & C. Lockwood (Eds), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions. Version 1 (updated August 2011). Cochrane Collaboration Qualitative Methods Group*. Retrieved from <http://cqrng.cochrane.org/supplemental-handbook-guidance>
- Harden, A., Garcia, J., Oliver, S., Rees, R., Shepherd, J., Brunton, G., & Oakley, A. (2004). Applying systematic review methods to studies of people's views: an example from public health research. *Journal of Epidemiology and Community Health*, 58, 794-800. doi:10.1136/jech.2003.014829
- Hare, C., Davis, C., & Shepherd, M. (2006). Safer medicine administration through the use of e-learning. *Nursing Times*, 102, 25-27. Retrieved from <http://www.nursingtimes.net/>
- Health Care Practice Research & Development Unit. (2009). *Evaluation tool for qualitative research*. Retrieved from http://usir.salford.ac.uk/12970/1/Evaluation_Tool_for_Qualitative_Studies.pdf
- Hurst, J. (2005). Evaluating staff and student experiences of multidisciplinary continuous professional development via distance-learning. *EDTNA/ERCA Journal*, 31, 160-163. Retrieved from <http://www.edtnaerca.org/pages/education/jrc.php>
- Kinghorn, S. (2005). Delivering multiprofessional web-based psychosocial education – the lessons learnt. *International Journal of Palliative Nursing*, 11, 432-437. Retrieved from <http://www.ijpn.co.uk/>
- Kirby, D., Brener, N., Brown, N., Peterfreund, N., Hillard, P., & Harrist, R. (1999). The impact of condom availability [correction of distribution] in Seattle schools on sexual behavior and condom use. *American Journal of Public Health*, 89, 182-187. doi: 10.2105/AJPH.89.2.18

- Larsen, T., & Jenkins, L. (2005). *Evaluation of online learning module about sickness certification for general practitioners*. Retrieved from <http://193.129.121.133/asd/asd5/rports2005-2006/rrep304.pdf>
- Lincoln, Y. (1995). Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry*, 1, 275-289. doi: 10.1177/107780049500100301
- Mays, N., & Pope, C. (1995). Qualitative research: Rigour and qualitative research. *British Medical Journal*, 311, 109-112. doi: 10.1136/bmj.311.6997.109
- Miles, M., & Huberman, A. (1994). *Qualitative Data Analysis: A Sourcebook of New Methods*. (2nd ed.) Newbury Park, CA: Sage.
- Nelson, M., & Quinney, D. (1997). Evaluating a school-based health clinic. *Health Visitor*, 70, 419-421.
- Noyes, J., & Popay, J. (2007). Directly observed therapy and tuberculosis: how can a systematic review of qualitative research contribute to improving services? A qualitative meta-synthesis. *Journal of Advanced Nursing*, 57, 227-243. doi: 10.1111/j.1365-2648.2006.04092.x
- Oliver, S., Rees, R., Clarke-Jones, L., Milne, R., Oakley, A., Gabbay, J., Stein, K., . . . Gyte G. (2008). A multidimensional conceptual framework for analysing public involvement in health services research. *Health Expectations*, 11, 72-84. doi: 10.1111/j.1369-7625.2007.00476.x
- Paterson, B., Thorne, S., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research. A practical guide to meta-analysis and meta-synthesis*. Thousand Oaks, CA: Sage.
- Patton, M. (1990). *Qualitative evaluation and research methods*. (2nd ed.) Newbury Park, CA: Sage.
- Pawson, R. (2006). *Evidence based policy: A realist perspective*. London: Sage.

- Petticrew, M., & Roberts, H. (2006). How to appraise the studies: An introduction to assessing study quality. In M. Petticrew & H. Roberts (Eds.), *Systematic reviews in the social sciences: A practical guide* (pp. 125-163). Oxford: Blackwell.
- Popay, J., Rogers, A., & Williams, G. (1998). Rationale and standards for the systematic review of qualitative literature. *Qualitative Health Research*, 8, 341-351. doi: 10.1177/104973239800800305
- Power, R. (2001). Checklists for improving rigour in qualitative research. Never mind the tail (checklist), check out the dog (research). *British Medical Journal*, 323, 514-515. doi: 10.1136/bmj.322.7294.1115
- Public Health Resource Unit, Critical Appraisal Skills Program. (2009). *10 questions to help you make sense of qualitative research*. Retrieved from <http://www.sph.nhs.uk/sph-files/casp-appraisal-tools/Qualitative%20Appraisal%20Tool.pdf>
- Sandelowski, M., & Barroso, J. (2007). *Handbook for Synthesizing Qualitative Research*. New York: Springer.
- Sandelowski, M., Barroso, J., & Voils, C. (2007). Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Research in Nursing and Health*, 30, 99-111. doi: 10.1002/nur.20176
- Schulz, K., Altman, D., & Moher, D. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, 340, c332. doi: 10.1136/bmj.c332
- Schuster, M., Bell, R., Berry, S., & Kanouse, D. (1997). Students' acquisition and use of school condoms in a high school condom availability program. *Pediatrics*, 104, 689-694. doi: 10.1542/peds.100.4.689

- Tanner, K., Kirton, A., Stone, N., & Ingham, R. (2003). *Evaluating the effectiveness of Time 4U services and associated educational interventions in Worcestershire. Summary of main findings*. Worcester, UK.
- Thomas., J., & Harden., A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45. doi: 10.1186/1471-2288-8-45
- Thorley, K., Turner, S., Hussey, L., Hall, N., & Agius, R. (2007). CPD for GPs using the THOR-GP website. *Occupational Medicine*, 57, 575-580. doi: 10.1093/occmed/kqm116
- Thorne, S. (2011). Toward Methodological Emancipation in Applied Health Research. *Qualitative Health Research*, 21, 443-453. doi:10.1177/1049732310392595
- Washkansky, G. (2008). *Sexual health drop in service*. London: London Borough of Hammersmith & Fulham.
- Whittemore, R., Chase, S. K., & Mandle, C. (2001). Validity in qualitative research. *Qualitative Health Research*, 11, 522-537. doi: 10.1177/104973201129119299
- Whittington, K., Cook, J., Barratt, C., & Jenkins, J. (2004). Can the Internet widen participation in reproductive medicine education for professionals? *Human Reproduction*, 19, 1800-1805. doi: 10.1093/humrep/deh333
- Wilkinson, A., Forbes, A., Bloomfield, J., & Fincham, G. (2004). An exploration of four web-based open and flexible learning modules in post-registration nurse education. *International Journal of Nursing Studies*, 41, 411-424. doi:10.1016/j.ijnurstu.2003.11.001
- Zabin, L., Stark, H., & Emerson, R. (1991). Reasons for delay in contraceptive clinic utilization. Adolescent clinic and nonclinic populations compared. *Journal of Adolescent Health*, 12, 225-232. doi:10.1016/0197-0070(91)90015-E

Zeanah, P., Morse, E., Simon, P., Stock, M., Pratt, J., & Sterne, S. (1996). Community reactions to reproductive health care at three school-based clinics in Louisiana. *Journal of School Health*, 66, 237-241. doi: 10.1111/j.1746-1561.1996.tb06277.x

Bios

Christopher Carroll, PhD, is a senior lecturer in health technology assessment at the School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom.

Andrew Booth, MA, is a reader in evidence based information practice and director of information at the School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom.

Myfanwy Lloyd-Jones, DPhil, is a senior research fellow at the School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom.

Table 1. Reporting Assessment Checklist

Criteria	Categorization	Definition
The question and study design	Yes	If the choice of study design was given and explained
	No	If article does not specify question and study design
The selection of participants	Yes	If the selection of participants is described explicitly as e.g., purposive, convenience, theoretical and so forth.
	No	If only details of participants are given
Methods of data collection	Yes	If details of the data collection method are given e.g., piloting; topic guides for interviews; number of items in a survey; use of open or closed items; validation, and so forth.
	No	If just only states focus group, interview or questionnaire
Methods of analysis	Yes	If details of analysis method are given, e.g., transcription and form of analysis (with reference or full description of method), validation tests, and so forth.
	No	If only states content analysis or that data were analyzed

Figure 1 Illustrative examples of studies assessed as adequately and inadequately reported

		Adequately reported		Inadequately reported	
	Study Identifier	Salmon & Ingram (2008)	Yes/No/Unclear	Tanner et al (2003)	Yes/No/Unclear
Criterion	Study design & question	"The focus groups and interviews offered the opportunity to explore barriers to attendance and strategies that from the young people's perspective would support future engagement."	Y	Not reported	N
	Participant selection	"In three schools, young people who potentially had not attended the service were asked to participate in small focus group discussions . . . or one to one interviews . . . attention paid to involving boys, hard to reach groups and those who may be involved in early sexual activity . . . For ethical reasons it was inappropriate to single out those not using the service. The groups therefore included ...".	Y	Not reported	N
	Data collection	Focus groups and interviews (no additional details)	N	Interviews	N
	Analysis	"During data analysis interview transcripts were analyzed using the recognized qualitative data analysis approach of sorting quotations from the transcripts into data units or themes and subthemes. This was done using . . . In particular, they focused on the barriers and reasons young people may have for not accessing the service. Respondent validation (Silverman 2000), whereby interview scripts or aspects of the analysis are returned . . . they contributed to the evaluation and service development."	Y	"The information from the interviews was processed using horizontal and vertical thematic analysis techniques to identify both similarities and differences between the thoughts, attitudes and experiences of the interviewees . . .".	Y

Appendix E - Paper M4

© 2011 Carroll et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License

(<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Publishers' version:

Carroll C, Booth A, Cooper K. A worked example of "best fit" framework synthesis: a systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Med Res Methodol*. 2011 Mar 16;11:29.

<http://www.biomedcentral.com/1471-2288/11/29>

RESEARCH ARTICLE

Open Access

A worked example of “best fit” framework synthesis: A systematic review of views concerning the taking of some potential chemopreventive agents

Christopher Carroll*, Andrew Booth, Katy Cooper

Abstract

Background: A variety of different approaches to the synthesis of qualitative data are advocated in the literature. The aim of this paper is to describe the application of a pragmatic method of qualitative evidence synthesis and the lessons learned from adopting this “best fit” framework synthesis approach.

Methods: An evaluation of framework synthesis as an approach to the qualitative systematic review of evidence exploring the views of adults to the taking of potential agents within the context of the primary prevention of colorectal cancer.

Results: Twenty papers from North America, Australia, the UK and Europe met the criteria for inclusion. Fourteen themes were identified *a priori* from a related, existing conceptual model identified in the literature, which were then used to code the extracted data. Further analysis resulted in the generation of a more sophisticated model with additional themes. The synthesis required a combination of secondary framework and thematic analysis approaches and was conducted within a health technology assessment timeframe.

Conclusion: The novel and pragmatic “best fit” approach to framework synthesis developed and described here was found to be fit for purpose. Future research should seek to test further this approach to qualitative data synthesis.

Background

While the potential limitations of qualitative data synthesis are frequently articulated, so is the utility of conducting such analysis [1]. Framework synthesis is one of several methodologies currently being developed for synthesising qualitative data [2]. This type of synthesis is based on framework analysis [3] and “offers a highly structured approach to organising and analysing data (e.g. indexing using numerical codes, rearranging data into charts etc.)” [2]. It involves the preliminary identification of *a priori* themes against which to map data from included studies. In contrast to such methods as meta-ethnography [4], framework synthesis is primarily a deductive approach. As such it carries certain

pragmatic advantages which might prove beneficial within the constraints of a health technology assessment where effectiveness review, economic evaluation and qualitative evidence synthesis are conducted together within tight time constraints. Thus a framework may not simply be an instrument for analysis but may also represent a scaffold against which findings from the different components of an assessment may be brought together and organised. Limited numbers of published examples of “framework synthesis” exist, among which the most prominent have been produced by the same team at the Institute of Education, University of London [5-7]. The present synthesis therefore represents an early worked example of this approach, the only one originating from outside of the team who developed the method, and offers an opportunity for further methodological advances. It is also the first to explore the

* Correspondence: c.carroll@shef.ac.uk
Health Economics and Decision Science (HEDS), School of Health and Related Research (SchARR), University of Sheffield, Sheffield, UK

strengths and limitations of a pragmatic “best fit” approach using an existing conceptual model as a starting point to identify *a priori* themes.

This qualitative evidence synthesis was originally designed to complement a systematic review and economic evaluation on the prevention of colorectal cancer by reviewing evidence relating both to the attitudes of adults concerning the taking of named chemopreventive agents and factors that may inform the related, perceived risk-benefit balance [8]. The agents of interest were non-steroidal anti-inflammatory drugs (NSAIDs, including aspirin), vitamins, minerals, folic acid or folate, selenium, calcium and dietary supplements generally. No previous evidence synthesis was identified regarding people’s views about taking these agents, especially for primary prevention of colorectal cancer. The effectiveness of any agent is moderated by levels of compliance with the proposed regimes. For those contemplating taking such agents, for example to protect against cancer, the decision-making process can be seen as complex, due to the uncertainty of the “trade-off” between efficacy of the agent, i.e. the likelihood of getting the cancer, and its possible long-term side effects [9]. It has also been pointed out that people may find it difficult to incorporate a regular pattern of chemoprevention into the demands of day-to-day life. On the other hand research points to the successful use of low-dose aspirin in reducing the risk of heart attack and stroke [10].

The aim of the current paper is to summarise key results of this synthesis of qualitative studies within the context of describing the application of a “best fit” method, and to consider the lessons learned from adopting such an approach to framework synthesis.

Methods

Search methods

The aim of the qualitative evidence synthesis was to examine people’s attitudes towards the taking of agents or supplements that may be used in the primary prevention of colorectal cancer, i.e. NSAIDs (including aspirin), vitamins, minerals, folic acid or folate, selenium, calcium and dietary supplements generally. The synthesis included studies that focused on exploring the views, beliefs or attitudes of people who took any of these agents for any purpose. A systematic search to identify relevant studies was performed by an information specialist following piloting of appropriate search strategies. The search combined terms describing the agents of interest (NSAIDs, aspirin, vitamins, etc.) with a published, validated filter for identifying qualitative studies, together with the medical subject heading “qualitative research” [11]. The full search strategy is available in the **Appendix**. Databases searched for published and unpublished material included MEDLINE, PreMEDLINE,

CINAHL, EMBASE, AMED, ASSIA, IBSS, PsycINFO, Science Citation Index, and Social Science Citation Index, and the HMIC and King’s Fund databases. Studies were limited to those in English published from 2003 onwards to capture contemporary views and attitudes. Searches were undertaken in June 2008. Given the problems with identifying social science or qualitative literature through systematic searching of electronic databases alone [12,13], the reference lists of all included studies were checked for additional literature, and a “berry-picking approach” utilising supplementary, non-systematic searching [14] testing various combinations of terms was also performed by two of the authors (AB, KC). This iterative, pragmatic approach to searching aimed to identify a set of studies providing relevant information on views and attitudes towards the taking of potential chemopreventive agents.

Study selection

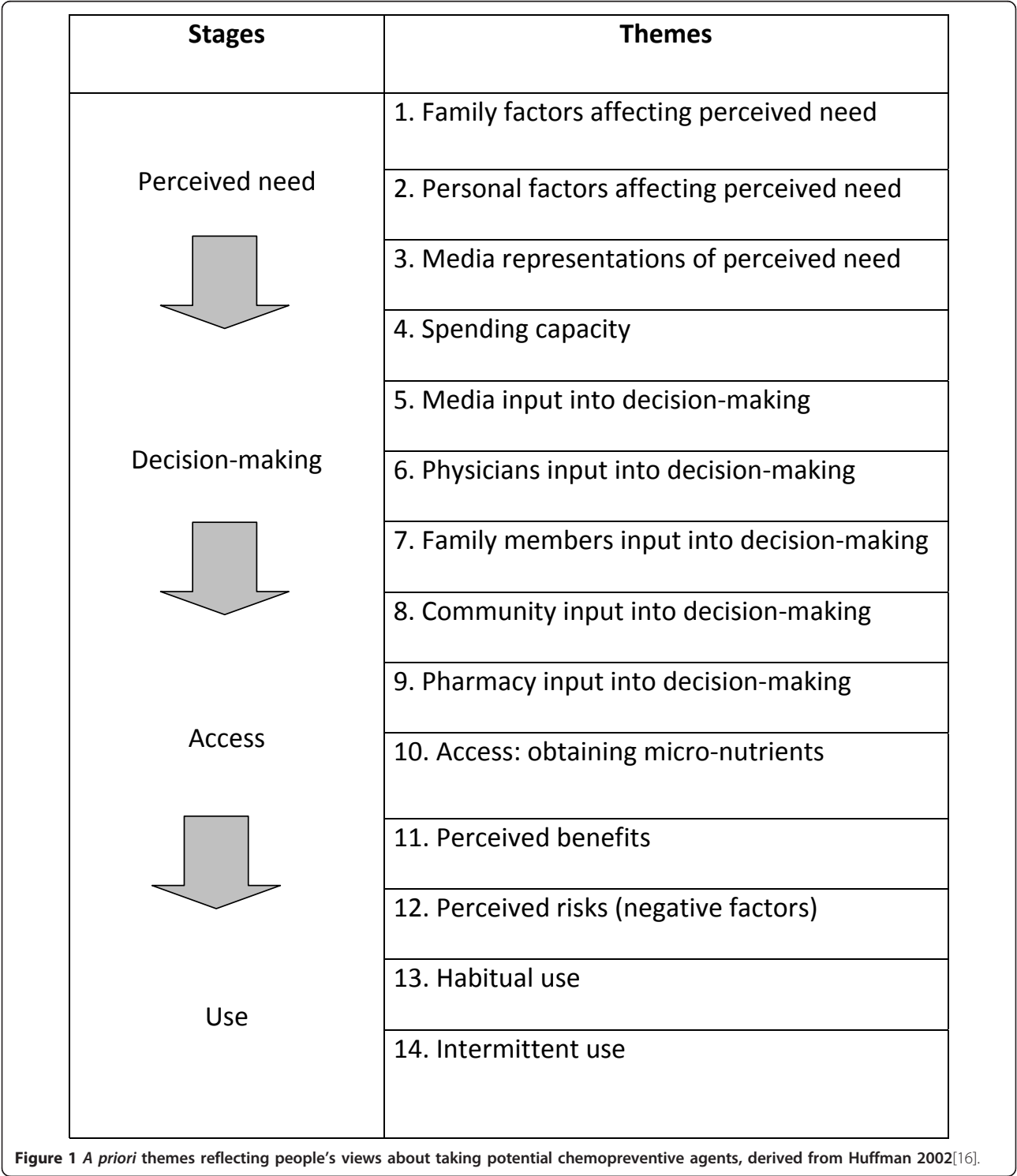
To be included in the review, a study had to focus on exploring the attitudes, perceptions and beliefs of adults (any country) surrounding the taking of the agents listed above, through qualitative data from interviews or focus groups, and cross-sectional data from satisfaction surveys, i.e. unstructured and structured, but often textual data describing people’s own, personal, subjective experiences, views or attitudes relating to the intervention of interest. Previous reviews have also adopted this inclusive approach to “views” studies, i.e. including qualitative data describing people’s attitudes and beliefs from satisfaction surveys as well as more traditional qualitative study designs [6,15]. The authors each screened a third of the citations for relevance (based on the inclusion criteria) and references for potential inclusion were discussed within the team. Disagreements or uncertain inclusions were resolved by discussion or by retrieval of the full paper to make a definitive judgment. Full papers of all potentially relevant citations were screened using the same process. Data from the included studies were extracted by two of the authors (CC, KC) using a review-specific form developed following piloting on one included paper.

“Best fit” approach to framework synthesis

The authors chose the framework synthesis approach because a published model was identified from the literature that conceptualised attitudes of adult women to the taking of vitamins and minerals [16]. The approach therefore was augmentative and deductive (building on this existing model or framework), rather than grounded or inductive (starting with a completely blank sheet). The model identified did not entirely match the topic under study, but it was a “best-fit” and provided a relevant pre-existing framework and themes against which

to map and code the data from the studies identified for this review. A list of themes was derived from this model (see Figure 1) and provided the *a priori* framework of themes against which to code the data extracted from the included studies.

Data for analysis consisted either of *verbatim* quotations from study participants or findings reported by authors that were clearly supported by study data, for example, ‘four of the five interviewees reported that the views of family and friends affected their decision-making’ or



'75% of respondents said that they were concerned about side effects of NSAIDs'. These data were extracted from the "Results" sections of included studies only, as it was felt that the Discussion and Conclusion sections would not present any new data, only additional interpretation or contextualisation of a study's 'findings'. Two of the authors (CC, KC) each extracted data from half of the included studies. Where any relevant data from the included studies did not translate into any pre-existing themes, a method was required to capture these data for the analysis. The published descriptions of framework synthesis do not specify a particular method for this, so the authors applied secondary thematic analysis, an interpretive, inductive approach grounded in the data based on methods from primary research, whereby additional themes were created as needed based on the study data [17]. In this way, the existing model acted as the basis for the synthesis and could be built-upon, expanded upon, reduced or added to by these new data. Each reviewer checked and examined critically the extraction and categorisation or coding of data performed by the other. The principal aim of this process was to examine the first reviewer's categorisation of the data, i.e. either to verify the coding or to challenge it by offering an alternative.

The authors then discussed the data and resulting themes, both those from the pre-existing model and those generated by the novel, inductive thematic analysis of the extracted study data. A consensus was reached on which *a priori* themes were supported by the data, and whether new themes identified by the reviewers did actually map either to a pre-existing theme or to one another (c.p. reciprocal translation [2]). The result was a finalised list of themes. The primary reviewer (CC) then offered an interpretation of the relationships between the themes based in part on the relationships as they were represented in the original model (see Figure 1), and also based on the data itself, which suggested, for example, that "the media" inputted into the central procedural themes of both perceived need and decision-making. The new model was then critically considered by all reviewers. A revised conceptual model was therefore developed building on the earlier, identified model, to describe and explain people's views around the taking of potential chemopreventive agents.

Consideration of study quality

Published descriptions of framework synthesis typically exclude studies of lower quality. However this was not the approach used in this case, representing a further innovative deviation from the published method [2]. All studies that satisfied the relevance criteria were included because there is an increasingly strong case for not excluding qualitative data studies from evidence synthesis based on quality assessment [1,18,19]. Studies were

assessed using key quality criteria derived from relevant critical appraisal checklists for qualitative studies [20] and other systematic reviews of people's views [1,2]. These elements also appear in recent guidance from the Cochrane Qualitative Research Methods Group [21]. The assessment consisted of querying whether the following are clearly and adequately described in the publication: the question and study design; how the participants were recruited or selected; and the methods of data collection and analysis used (See Additional file 1). The "better-reported" studies provided details on two or more criteria, whereas the "inadequately-reported" studies clearly described no more than one. The decision only to focus on these four elements, and what was reported or clearly described by the included studies, was taken for two reasons. Firstly, these elements of the study were potentially more easily judged and apprehended than others, as they were either described or not. Secondly, it has been pointed-out previously that any appraisal checklist is only assessing what has been reported in a publication [22]. The focus therefore was on the reporting of basic methods and not potentially subjective judgements regarding studies' validity or reliability [18].

While it is acknowledged that there is always uncertainty concerning how well or poorly a study has been conducted, if authors clearly describe their approach and sampling, and data collection and analysis methods, then this potentially lends greater robustness to the study's findings. This is because any inherent "risk of bias" may be better determined than if this information was absent, regardless of the study's findings. This does not preclude the possibility that an "inadequately-reported" study has actually been well-conducted, but it does form a reasonable basis for making a quality assessment. This relatively small number of easily-defined criteria can also be seen to apply to qualitative studies universally and may be more practical than checklists with much larger numbers of questions, especially as these have been found to generate low inter-rater reliability scores among otherwise experienced qualitative systematic reviewers [18]. This was one of the first practical attempts to utilise assessment criteria based specifically and exclusively on the description or reporting of a study's method and sampling strategies, and methods of data collection and analysis. No study was excluded on the basis of the adequacy of its reported processes, but the assessment aimed to explore quality of reporting as an explanation for differences in the results of otherwise similar studies, and to consider its impact on the internal validity of the review [23]. A sensitivity analysis would be performed in the event of the inclusion of "inadequately-reported" studies.

Results

Quantity and quality of included studies

The literature search identified 1,805 unique citations, 15 of which satisfied the inclusion criteria. Five further studies were identified by the “berry picking” approach described above [8]. In total, twenty studies were included. No study failed to describe clearly at least two of the following: the question and study design, and the methods of sampling, data collection or analysis. Study quality, in terms of how well or how poorly studies were described, was therefore not a potential moderator of the findings; a sensitivity analysis was not performed.

Data synthesis and development of model

A combination of coding against pre-existing themes and the generation of and assignment of data to new, agreed themes, generated the model presented in Figure 2. A full description of the evidence supporting this model is published elsewhere [8]. The model describes the processes involved in an individual’s decision about whether or not to take possible chemopreventive agents. The process runs from the first stages of perceived need, on the left, through the decision-making process itself, to final non-use or use, and maintenance of use, on the right. External

agents, such as health professionals and family members, and internal factors, such as a person’s own experience or health, were all found to impact both on an individual’s perceived need for an agent or supplement, as well as their subsequent decision about whether or not to take it.

Usefulness of the preliminary conceptual framework in assigning data to themes

Since the source of the preliminary framework was a single published model, the manner in which new themes built-on, developed and altered this preliminary conceptual framework is quite transparent. In this review, this may be assessed in part by comparing Figure 1 with Figure 2. The principal procedural elements of the preliminary model also held true for this sample of studies and their population, i.e. the transition through the stages of perceived need, decision-making, risk versus benefit and use or non-use. These elements also reflect the three key stages of Contemplation, Determination and Action in Prochaska and Velicer’s model (1997) of the development of health behaviours, which was later found to be relevant [24]. The *a priori* identification of these key constructs therefore enabled

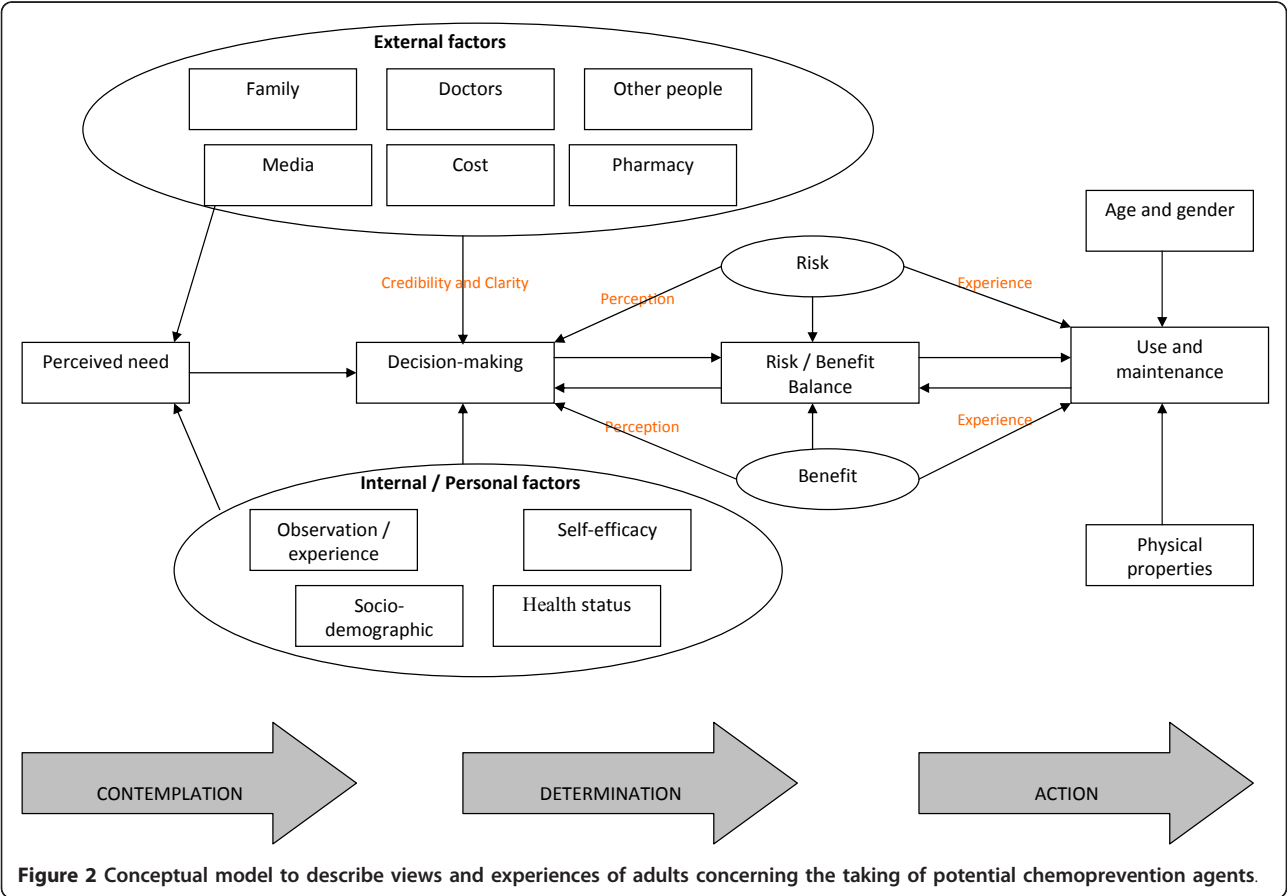


Figure 2 Conceptual model to describe views and experiences of adults concerning the taking of potential chemoprevention agents.

the rapid coding of study data from this review against these tested and highly relevant components of health behaviour decision-making. The preliminary framework also provided “themes” that informed the “perceived need” and “decision-making” stages of the model (see numbers 1-9 in Figure 1). Once clear definitions had been applied to each of these themes, the study data were coded rapidly against them. Very little study data were coded against the themes of “Spending capacity” (or “Costs”) and “Access: obtaining the agent”, which may reflect differences in the cultural context of the preliminary conceptual model (a low-income country in South America) compared with the studies included in the review (principally UK, Europe and North America). However, relatively more substantial amounts of data were coded against the remaining themes.

Extension of the preliminary conceptual framework to generate the final model

Despite these helpful overlaps, which permitted rapid and reliable coding of much data from the included studies, the preliminary model lacked sufficient depth or complexity to explain all the data in the included studies. As with the preliminary model, some factors influenced both need and decision-making. For example, the influences of family and the media were present at both of these stages, but the categorisation of these factors was re-specified in the new model. Family, media, physicians, other people and pharmacy were all designated in the new model as external factors having input into perceived need and decision-making. The “personal factors” theme from the original model was re-specified as “internal or personal factors” to include an individual’s own observations or experience, their health and socio-economic status, age and gender, and their sense of self-efficacy. All of these characteristics were found in the included studies to affect perceived need and decision-making. It was felt that the pre-existing theme of “personal factors” alone was insufficient to illustrate the complexity of factors at play. The role of age, gender and the physical properties of agents were new factors identified by the synthesis affecting the *a priori* theme of use, which were absent from the original conceptual model.

Relationships between the themes were not well-developed in the preliminary model. The synthesis found that not only did family, physicians and others affect decision making, but also that this relationship was moderated by the credibility of the source and the clarity of the information being given. Perceived risks and benefits were key pre-existing themes shaping use, but the moderating role of personal experience was an additional element identified by the synthesis for the new model. Furthermore the risk/benefit balance theme was also found to

have an ongoing, potentially recursive influence on decision-making and agent use. Indeed, unlike the existing models, which appear to be exclusively linear, the model that resulted from this synthesis was potentially more recursive: the decision-making stage might still be revisited on the basis of side-effects (“risks”) experienced at the stage of use. This new model can therefore be seen not only to validate, but also to build upon, extend and contextualise existing, relevant published models. The *a priori* boxes of Contemplation, i.e. perceived need; Determination, i.e. decision-making; and Action, i.e. use and maintenance, have been opened to reveal the complexities of the factors therein, their relationships and moderators.

Discussion

The model generated by the framework synthesis describes the processes involved in an individual’s decision about whether to initiate and keep taking potential chemopreventive agents. External agents, such as health professionals and family members, and internal factors, such as a person’s own experience or health status, combine to impact on an individual’s perceived need for an agent or supplement, and their subsequent decision about whether or not to take it. Decision-making was strongly influenced by perceived risks and benefits associated with an agent or supplement. Firstly, perceived risks and benefits directly influence an individual’s decision to take an agent. Secondly, they may inform a personal assessment of the trade-off between risk and benefit, thus affecting the decision-making process. It has been reported elsewhere that decision-making regarding agents for chemoprevention or symptom management may be affected both by health status, for example, a cancer diagnosis [25,26], and by people’s perceived need for an agent and perceived risks associated with that agent [27-29]. The model generated by this review highlights the complex influences at work in this decision-making process.

This review applied a form of framework synthesis to analyse the data, based on a single “best fit” model identified in the literature. This approach differs from other published versions of framework synthesis in which the *a priori* framework was developed from a range of sources, including familiarisation with and consultation around the published background literature, both theoretical and empirical, and personal experiences [5,6]. The approach taken here is of potential value for systematic reviewers as it does not require such extensive literature review, consultations or topic expertise to develop an *a priori* framework before embarking on the review itself. This may be of particular value when undertaking a synthesis of qualitative evidence within the limited timeframes of a health technology

assessment, for example. Projects such as Health Technology Assessments, produced in multidisciplinary centres with contractual obligations, with a six-month or one-year span, and which also involve reviews of effectiveness, cost-effectiveness, mathematical modelling and, in some cases, qualitative evidence synthesis, often present challenges in relation to timeliness and the availability and expertise of members of research teams [30]. In this particular case study, the qualitative evidence synthesis was conducted after the effectiveness synthesis, which required the qualitative synthesis to be fairly quick within the project's required timeframe. However, a temporal dependency between the two types of synthesis will not always exist, and so a more in-depth qualitative approach may be possible for some projects. However, if a framework of related, relevant concepts already exists, then the approach used here permits a far more rapid identification of the *a priori* framework; it also permits more rapid and structured coding and synthesis of data from the review's included studies than grounded-theory techniques. In this way, where existing theories or models exist, they can be tested against the evidence for the review's own particular criteria and evidence. This approach is therefore potentially more pragmatic than other forms of qualitative data synthesis. The identification and use of a model that was overtly "best fit", and therefore carried shared acknowledgment within the team that it was contingent on emerging data also empowered the reviewers to resist the inclination to "slot" study findings into a generic framework. This potentially enabled individual team members to privilege context-specific insights that emerged from this review over the generic observations already present within the pre-existing model. Furthermore it provided a mechanism for flagging up and explicitly communicating divergent findings or themes within the review team. The resultant synthetic product is expressed as an enhanced model recording each key dimension identified; the nature of the concepts under study; and associations between themes and tensions between them [6].

The method is however dependent on the identification of an appropriate existing conceptual model. The review team sought to identify such a model by combining a sensitive string of search terms (e.g. model\$ OR framework\$ OR theoretical OR theory OR concept OR conceptual) with terms representing the health-related behaviour of interest. This approach was employed firstly on a bibliographic database (PubMed MEDLINE) but was found to be limited by poor coverage of theoretical aspects in published abstracts. A more productive approach proved to be using Google Scholar with the same string of search terms, and certainly the potential for this approach to be used with other collections of full-text documents remains to be further explored. This

strategy was conceived as iterative and purposive: it required search strategies that aimed to maximise the likelihood of retrieving a model of pragmatic utility to the project; the aim was not the systematic identification of all such models.

Furthermore the approach used for this particular case study was predicated on the review team's belief that the key criterion of the appropriateness of such a model most likely related to the health-related behaviour of interest, i.e. attitudes to the long-term taking of particular dietary supplements or similar agents. The population and the agents themselves may be less critical in such cases, although the closer the fit to the population and intervention of interest, the better. This is why we describe it as a "best-fit" approach. In this case study, young women and vitamins or micro-nutrients formed a sub-set of the populations and agents of interest. The conceptual model therefore had limited external validity but was still externally valid.

Some issues were encountered when piloting this "best fit" framework synthesis method. When initially seeking to code the extracted data from the included studies using the themes derived from the relevant model, the two reviewers were not always coding the same data against the same themes. It therefore became apparent that each of the *a priori* themes had to be clearly defined in order to facilitate the coding process. The subsequent provision of clear consensual definitions not only enhanced the reliability of the coding, but also strengthened the rigour of the synthesis. It should be recognised, however, that while consensus between reviewers strengthens internal validity this does not necessarily ensure congruence with the original meanings intended by the author of the framework (external validity). In this sense a form of "reciprocal translation" is taking place but via use of a conceptually rich "index paper" (many-to-one), rather than across all included studies (many-to-many), as intended by the originators of meta-ethnography [31]. Such considerations have been neither identified nor articulated in previous studies.

It further became apparent that additional analysis was needed to interpret and analyse data which could not be reliably assigned to any of the pre-existing, *a priori* themes, or, in the case of "personal factors", for which the pre-existing theme was inadequate. In this sense the usefulness of a particular framework is not only determined by "conceptual fit" but also by pragmatic concerns of what proportion of the study data can be accommodated within it. Further thematic analysis of data from the included studies was therefore required. This was completed by the first author using standard thematic analysis techniques, and the results examined critically by the other two reviewers. The resulting,

agreed new themes were then incorporated with the pre-existing themes into a new conceptual model that captured the data and reflected a possible network of relationships between those data-driven themes. The existing published descriptions of the framework synthesis method do not detail particular techniques for analysing data that are not captured by the preliminary framework, how any such new themes are to be incorporated into the final model, or how the relationships between these themes may be expressed.

Finally, this review did not exclude studies on the basis of quality, thereby deviating from one element of the published description of framework synthesis [2]. The internal validity of a review depends in part on the quality of included studies and the reliability of their findings. Currently there is much debate and little consensus around the feasibility and usefulness of quality assessments of qualitative studies in evidence synthesis [18]. Some techniques, such as meta-ethnography [4], and the previously published form of framework synthesis, actively exclude studies on the basis of the quality assessment. The quality assessment for this review focused on reporting of study design, sampling strategies and methods used for data collection and analysis. These items were the most frequently reported and easily apprehended elements of study design. They thus offered a reasonable route for identification of potential risk of bias. All twenty included studies were assessed as being of similar, generally satisfactory "quality", so, from this perspective, study quality did not provide a potential explanation for any differences in findings. The issue of the inclusion or exclusion of studies for this type of synthesis, based on their assessed quality, therefore remains unresolved based on this case study.

Methodologically the authors found this "best fit" approach to framework synthesis, as developed and tested in this review, to be a useful, fairly rapid and reliable and, above all, pragmatic method of synthesising qualitative data. This "best fit" approach to synthesis was therefore found to work well overall, particularly within the role previously identified as an existing strength, namely for testing existing potentially generalisable theories and models within a specific context. However, such a "best fit" approach would benefit from further testing and refinement.

Limitations

This is a single case study evaluating the approach described; additional studies testing this approach to qualitative evidence synthesis need to be undertaken. Also, as an approach, it is only viable if an appropriate model already exists in the literature. The other published models for framework synthesis circumvent this problem as the *a priori* framework is generated by the

research team itself. It is also the case that an apparently appropriate *a priori* model may be found only to accommodate a small proportion of the data from a review's included studies. In such a case, secondary thematic analysis would form the principal approach to synthesis, thus reducing the major potential pragmatic benefits of the best-fit approach described in this paper. Reviewers must therefore exercise careful consideration of the potential external validity of existing models based on the behaviour and population of interest.

Conclusion

This "best fit" method of framework synthesis utilised current methodological developments within qualitative data synthesis for systematic review and the production of accompanying conceptual models and frameworks. The case study was a systematic review of adults' views about taking various potential chemopreventive agents. The "best fit" framework synthesis offered a means to reinforce, critique and develop an existing published model, conceived for a different but relevant population. Being able to start from *a priori* themes, rather than generating theory grounded in data, produced a relatively rapid process when compared to more interpretative forms of synthesis. However this "best fit" method still requires analysis of data that are not captured by the preliminary model. The authors suggest that this "best fit" approach occupies a pragmatic middle ground between grounded theory-type and framework based syntheses and acknowledge the need for further evaluation.

Appendix

Database: CINAHL - Cumulative Index to Nursing & Allied Health Literature

Search Strategy:

- 1 vitamin\$.tw.
- 2 mineral\$.tw.
- 3 folate\$.tw.
- 4 selenium.tw.
- 5 calcium.tw.
- 6 exp Dietary Supplements/
- 7 Dietary Supplementation/
- 8 dietary supplement\$.tw.
- 9 non-steroidal\$.tw.
- 10 non steroidal\$.tw.
- 11 nonsteroidal\$.tw.
- 12 NSAID\$.tw.
- 13 antiinflammator\$.tw.
- 14 anti-inflammator\$.tw.
- 15 anti inflammator\$.tw.
- 16 aspirin\$.tw.
- 17 or/1-16
- 18 interview\$.tw.

19 experience\$.tw.
20 qualitative\$.tw.
21 exp Qualitative Studies/
22 or/18-21
23 17 and 22
24 limit 23 to yr="2003 - 2008"

Additional material

Additional file 1: The question and study design; how the participants were recruited or selected; and the methods of data collection and analysis used.

Acknowledgements

The case study on which this work was based was part of a larger project funded by the UK NCCHTA (06/70/01)

Authors' contributions

CC and AB conceived the study; CC designed the study; CC, KC and AB extracted the data and appraised included studies; CC, KC, and AB analysed and interpreted the data. CC drafted the paper and KC and AB undertook critical revision of important content of the manuscript. All authors approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 13 December 2010 Accepted: 16 March 2011

Published: 16 March 2011

References

1. Thomas J, Harden A: Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology* 2008, **8**.
2. Barnett-Page E, Thomas J: Methods for the synthesis of qualitative research: a critical review. *BMC Medical Research Methodology* 2009, **9**.
3. Pope C, Ziebland S, Mays N: Qualitative research in health care: Analysing qualitative data. *British Medical Journal* 2000, **320**:114-116.
4. Campbell R, Pound P, Pope C, Britten N, Pill R, Morgan N, et al: Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care. *Social Science & Medicine* 2003, **65**:671-684.
5. Oliver S, Rees R, Clarke-Jones L, Milne R, Oakley A, Gabbay J, et al: A multidimensional conceptual framework for analysing public involvement in health services research. *Health Expectations* 2008, **11**:72-84.
6. Brunton G, Oliver S, Oliver K, Lorenc T: A Synthesis of Research Addressing Children's, Young People's and Parents' Views of Walking and Cycling for Transport London. London, EPPI-Centre, Social Science Research Unit, Institute of Education, University of London; 2006.
7. Nilsen E, Myrhaug H, Johansen M, Oliver S, Oxman A: Methods of consumer involvement in developing healthcare policy and research, clinical practice guidelines and patient information material. *Cochrane Database of Systematic Reviews* 2006, **3**: CD00456.
8. Cooper K, Squires H, Carroll C, Papaioannou D, Booth A, Logan R, Hind D, Tappenden P: Chemoprevention of colorectal cancer: systematic review and economic evaluation. *Health Technology Assessment* 2010, **14**:32.
9. Heisey R, Pimlott N, Clemons M, Cummings S, Drummond N: Women's views on chemoprevention of breast cancer: qualitative study. *Canadian Family Physician* 2006, **52**:624-625.
10. Lynch P: The role of aspirin in the prevention of polyp recurrence: What is the right dose? *Current Colorectal Cancer Reports* 2007, **3**:24-28.
11. Wilczynski N, Marks S, Haynes R: Search Strategies for Identifying Qualitative Studies in CINAHL. *Qualitative Health Research* 2007, **17**:705-710.
12. Papaioannou D, Carroll C, Booth A, Sutton A, Wong R: Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information and Libraries Journal* 2010, **27**:114-122.
13. Grayson L, Gomersall A: A difficult business: finding the evidence for social science reviews. *ESRC UK Centre for Evidence Based Policy and Practice. Working paper 19* 2003.
14. Sandelowski M, Barroso J: Searching for and Retrieving Qualitative Research Reports. In *Handbook for synthesizing qualitative research*. Edited by: Sandelowski M, Barroso J. New York, NY: Springer; 2007:35-74.
15. Thomas J, Harden A, Oakley A, Oliver S, Sutcliffe K, Rees R: Integrating qualitative research with trials in systematic reviews. *British Medical Journal* 2005, **328**:1010-1012.
16. Huffman S: Can Marketing Of Multiple Vitamin/Mineral Supplements Reach the Poor? The Vitaldia Project, Bolivia; 2002 [http://www.phishare.org/files/338_Can%20Marketing..Vitaldia.pdf].
17. Miles M, Huberman A: *Qualitative Data Analysis: A Sourcebook of New Methods*. 2 edition. Newbury Park, CA: Sage; 1994.
18. Dixon-Woods M, Sutton A, Shaw R, Miller T, Smith J, Young B, et al: Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *Journal of Health Services Research and Policy* 2007, **12**:42-47.
19. Dixon-Woods M, Bonas S, Booth A, Jones D, Miller T, Sutton A, et al: How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research* 2006, **6**:27-44.
20. Tong A, Sainsbury P, Craig J: Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care* 2007, **19**:349-357.
21. Hannes K: Critical appraisal of qualitative research. *Cochrane Qualitative Research Methods Group Handbook* 2010.
22. Dixon-Woods M, Shaw R, Agarwal S, Smith J: The problem of appraising qualitative research. *Quality and Safety in Health Care* 2004, **13**:223-225.
23. Golafshani N: Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report* 2003, **8**:597-607.
24. Prochaska J, Velicer W: The transtheoretical model of health behavior change. *American Journal of Health Promotion* 1997, **12**:38-48.
25. Demark-Wahnefried W, Peterson B, McBride C, Lipkus I, Clipp E: Current health behaviors and readiness to pursue life-style changes among men and women diagnosed with early stage prostate and breast carcinomas. *Cancer* 2000, **88**:674-684.
26. Strecher V, Rosenstock I: The health belief model. In *Health Behavior and Health Education*. Edited by: Rimer BK. San Francisco: Jossey-Bass; 1997:41-59.
27. Floyd D, Prentice-Dunn S, Rogers R: A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology* 2000, **30**:407-429.
28. Milne S, Sheeran P, Orbell S: Prediction and intervention in health-related behavior: a meta-analytic review of protection motivation theory. *Journal of Applied Social Psychology* 2000, **30**:106-143.
29. Horne R, Weinman J: Self-regulation and self-management in asthma: Exploring the role of illness perceptions and treatment beliefs in explaining non-adherence to preventer medication. *Psychology & Health* 2002, **17**:17-32.
30. Rotstein D, Laupacis A: Differences between systematic reviews and health technology assessments: a trade-off between the ideals of scientific rigor and the realities of policy making. *International Journal of Technology Assessment in Health Care* 2004, **20**:177-183.
31. Noblit GW, Hare R: *Meta-ethnography: synthesizing qualitative studies* Newbury Park, CA: Sage; 1998.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2288/11/29/prepub

doi:10.1186/1471-2288-11-29

Cite this article as: Carroll et al.: A worked example of "best fit" framework synthesis: A systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Medical Research Methodology* 2011 **11**:29.

Appendix F - Paper M5

**This is the Post-print version with changes from referees' comments.
Embargoed until October 2013. Publisher copyright and the source are
acknowledged as follows:**

**Booth A, Carroll C, Iltott I, Low LL & Cooper K. Desperately Seeking Dissonance:
Identifying the "Disconfirming Case" in Qualitative Evidence Synthesis.
Qualitative Health Research 2013; 23 (1): 126-141.**

Desperately Seeking Dissonance: Identifying the “Disconfirming Case” in Qualitative Evidence Synthesis

Andrew Booth, Chris Carroll, Irene Ilott, Lee Lan Low and Katy Cooper

Authors' Note: This study integrates methodological findings from two nationally funded projects from the National Institute for Health Research namely *the Contribution of nurses, midwives, and health visitors to protocol-based care and its variants* [Principal Investigator: Professor Malcolm Patterson] and *Chemoprevention of colorectal cancer: systematic review and economic evaluation* [Principal Investigator: Mr Paul Tappenden] and one from the Higher Education Academy, *Enhancing the student experience of workplace-based e-learning: a systematic review and best practice framework* [Principal Investigator: Mr Andrew Booth].

Correspondence concerning this article should be addressed to Andrew Booth, School of Health and Related Research (SchARR), University of Sheffield, Sheffield, South Yorkshire UK S1 4DA, e-mail: a.booth@sheffield.ac.uk.

Andrew Booth, MSc is a Reader in Evidence Based Information Practice at the School of Health and Related Research, University of Sheffield. His research interests include qualitative evidence synthesis, systematic review methodology and evidence based practice.

Christopher Carroll, PhD is a Senior Lecturer at the School of Health and Related Research, University of Sheffield. His research interests include qualitative evidence synthesis and the development and evaluation of systematic review methods more generally.

Irene Ilott, PhD is CLAHRC Knowledge Translation Project Lead at Sheffield Teaching Hospitals NHS Foundation Trust. Her research interests include evidence based practice, knowledge translation for sustaining change to benefit patient, staff and organizational outcomes, research capacity and workforce development.

Lee Lan Low, MA (Med Anthropology) is Research Officer at the Institute for Health Systems Research, Ministry of Health, Malaysia. Her research interests include qualitative research, health systems research methodology and knowledge translation

Katy Cooper, PhD is a Research Fellow at the School of Health and Related Research, University of Sheffield. Her research interests include systematic review methodology and reviewing of complex interventions.

Abstract

Actively seeking the disconfirming or deviant case is properly regarded as a hallmark of trustworthiness in primary qualitative research. The need to subject emergent theory to such testing is no less important within qualitative systematic reviews. There is, as yet, little available guidance on how to implement such strategies. Few researchers describe the practicalities of seeking the disconfirming case. We survey the methodological literature to gain a better understanding of how systematic reviews of qualitative research handle the disconfirming case. We reflect on our own experience from three recent qualitative evidence syntheses. We describe how reviewers might actively manufacture opportunities to identify discrepant or refutational findings. We conclude by outlining possible methods by which a team might integrate active seeking of a disconfirming case within the overall review process.

Keywords

meta-ethnography; metasynthesis; systematic reviews; triangulation

Qualitative research includes “the considered selection of typical, deviant, critical, or otherwise exemplary information rich cases” (Patton, 1990, p. 169). This requirement is shared by analysis of qualitative primary data from informants and by secondary analyses such as qualitative evidence syntheses of multiple study reports (including the specific sub-category of qualitative systematic reviews). “Deviant cases” are particularly important in the context of synthesis because, as its etymology suggests, reviewers are naturally inclined to look for commonalities or similarities (Petticrew & Roberts, 2006):

It is well known . . . that we select, evaluate, and remember information in a way that supports our individual preferences, we fail to look for evidence that disconfirms our pet hypotheses, and we cannot spot errors in our own reasoning (p. 130).

This quest for congruity, in preference to dissonance, is most clearly encapsulated in the averaging effect sought for quantitative studies through formal meta-analysis (Dong et al., 2008).

In primary research researchers identify the disconfirming case only after they have completed initial data collection and analysis. Identification of the disconfirming case involves selecting cases that serve as examples that do not fit emergent patterns (Patton, 1990). It requires strategies that complement and extend the team’s initial sampling strategies. Such cases allow the research team to evaluate rival explanations thereby enabling them to understand and define the limitations of findings from their own research. By seeking out disconfirming cases researchers are able to develop a richer, more in depth understanding of a phenomenon thus lending credibility to the resultant research account. In comparison to the averaging effect so typical of quantitative syntheses, identification of the “exception to the rule” requires that reviewers engage more extensively with the data to explore both complexity and contradictions.

Researchers often use the term deviant case, almost synonymously, as an alternative to the disconfirming case. Typically use of this phrase is associated with deviant case analysis by which researchers revisit and extend a theory to accommodate and interpret the findings from the deviant case.

In the context of an evidence synthesis the disconfirming case may relate to a specific subgroup of the population for whom an intervention that works on average, or for whom an intervention is broadly acceptable, does not achieve its expected impact. For example a universal telephone reminder system, which in theory should improve attendance at appointments, may not work for adolescents because they frequently change their mobile phone number. Alternatively the disconfirming case may represent a population for whom an intervention or programme works better than expected. So the same reminder system may prove particularly effective for a chronic population who are housebound and therefore always present to receive the telephone call. Such heterogeneity deserves explanation and investigation. At a theoretical level a particular aspect, previously considered significant, may no longer be considered important, perhaps because it is addressed implicitly by an intervention. A reviewer needs to explore and explain why this aspect, either through being completely absent or through being non-visible, does not figure in a published account. Of course such heterogeneity is not only present in differences between populations. It may equally appear in contextual or environmental differences or in preferences for different outcomes.

Although there is not an empirical base for recommendations regarding the optimal number of studies that a reviewer might include in a qualitative evidence synthesis one very important reason for limiting quantities of reports is to reduce the risk of the synthesis and

subsequent analysis being considered too superficial (Paterson et al., 2001). Indeed qualitative evidence syntheses that incorporate large numbers of study reports are often characterized as descriptive and aggregative rather than analytical and interpretative (Dixon-Woods et al., 2006a). For interpretative reviews there is a danger, particularly when synthesizing large numbers of studies, that a reviewer is more likely to miss the opportunities presented for identification of the disconfirming case.

Published guidelines for systematic reviews occasionally acknowledge the importance of identifying the disconfirming or deviant case. Nevertheless such guidelines provide relatively little guidance on how a review team might achieve such a goal (Noyes et al., 2008). In this report we seek to address this methodological gap. We undertook a brief literature survey of published examples of qualitative evidence syntheses, identified relevant methodological articles and drew on three examples from nationally-funded research in the United Kingdom. The primary aim was to find methods to maximize the possibility of identifying disconfirmatory findings and then to propose a toolkit from which to select to proceduralize this process. Therefore the purpose of this article is to provide an overview of how the disconfirming case has been handled in the meta-synthesis literature and to inform discussion of this issue. By doing this we hope to help researchers to select appropriately from a range of options according to the specific objectives, purposes and constraints of an individual project. The intention is not to specify a single template by which reviewers should pursue such findings. Indeed within primary research, identification of a disconfirming case is a strategy that researchers often use in iterative theory-building, theory-testing designs or data-driven research. Such a strategy is likely to prove of most value in evidence syntheses that share these overarching objectives. We outline fourteen such options (including strategies relating variously to the review team, the review methods, the

use of theory, and the presentation of findings). We describe our experience from three UK-funded evidence synthesis projects, all with significant qualitative components. We conclude with observations on the usefulness of different procedures and the respective contribution made by each procedure.

Methods

The Disconfirming Case in Qualitative Evidence Syntheses: a literature survey

To gain a better understanding of how systematic reviews of qualitative research have handled the disconfirming case, we conducted a brief literature survey of published examples of qualitative evidence syntheses and relevant methodological articles. We accessed a register maintained by the Cochrane Collaboration Qualitative Research Methods Group of examples of qualitative systematic reviews and methodological discussions, housed on a reference management database. This database contains references to over 300 methodological papers, 250 confirmed examples of published qualitative syntheses, and an additional 1200 references to potential syntheses. In addition, given that the concepts being explored are not typically present in titles or abstracts of such references, a Google Desktop tool was used to search an associated collection of the full text of many hundreds of articles in Portable Document Format and as Microsoft documents, both on the hosting computer and across the wider World Wide Web. All searches were conducted between July 2009 and January 2010. The team identified a variety of relevant terms. Three of these terms used the terminology of primary research (*deviant case*, *disconfirming case*, *negative case analysis*), and another was derived from a specific method of qualitative evidence synthesis (*refutational synthesis*).

Ten reviews reported identification of the disconfirming case (Paterson et al., 1998; Paterson et al., 1999; Thorne & Paterson, 1998; Thomas et al., 2003; Greenhalgh & Peacock,

2005; Dixon-Woods et al., 2006b; Sandelowski et al., 2007; Voils et al., 2008; O'Connell & Downe, 2009; Wong et al, 2010). Sixteen additional items discussed methodological considerations (Noblit & Hare, 1988; Jensen & Allen, 1996; Macintyre & Petticrew, 2000; Booth, 2001; Barroso & Sandelowski, 2003; Jones, 2004; Lloyd Jones, 2004; Pawson et al., 2004; Walker & Avant, 2005; Dixon Woods et al., 2006a; May, 2006; Petticrew & Roberts, 2006; Downe, 2007; Weed, 2007; Downe, 2008; Noyes et al., 2008). A process of “snowballing” was used to follow up appropriate references (Papaioannou et al., 2010). Snowballing is an umbrella term, analogous to techniques used in primary interview research, whereby identified respondents become the starting point for identification of additional participants. In the specific context of a literature search snowballing refers to using a known relevant item of literature to identify articles that have been cited, those that cite the article, related articles or subject terms used by the article as a starting point for inquiry. Snowballing was particularly appropriate given that identification of the disconfirming case is not typically the focus for study reports and might only be mentioned incidentally or in passing. Consequently this approach cannot be formally designated as a systematic review, simply a comprehensive search for methodological research to illuminate the chosen case studies.

A single reviewer read and re-read included studies. The same reviewer extracted data relating to dissonance, refutation or the deviant case to a single document, categorising and then coding extracts according to whether they related to characteristics of the team, characteristics of the method or to specific objectives of a review.

General approaches to the disconfirming case.

The meta-ethnography approach to evidence synthesis (Noblit & Hare, 1988) is one of the few methods to give explicit attention to identification of incongruities and inconsistencies. Indeed this approach embodies refutational synthesis as one of three methods for exploration of themes

within the data, alongside reciprocal translation and line of argument synthesis. Downe (2008)

comments:

Despite the frequency with which Noblit and Hare's 1988 work is cited, Noblit observed in 2004 that, while most published metasynthesis accounts describe reciprocal findings, few report on the refutational phase of the work. . . .(p. 7)

Writing within the meta-ethnography tradition, and as a precursor to development of the meta-study method, Paterson and colleagues (1998) acknowledge the potential value of diverse approaches to identifying and exploring dissonance in establishing the "trustworthiness" of the synthesis product:

Meta-ethnography in which trustworthiness was achieved by using multiple researchers, identifying negative or disconfirming cases, and testing rival hypotheses. (p. 57)

They continue by observing that "negative or disconfirming cases were identified when the data did not unanimously support the conclusions of the researcher" (p. 59). Booth (2001) identifies that one characteristic of systematic reviews of qualitative research is that "Particular attention is focused on negative or disconfirming cases. This adds to the richness of the insight that the review provides on the phenomenon of interest" (p. 2). Noticeably interpretative methods of synthesis, particularly those methods that trace their pedigree to grounded theory approaches, reflect a greater preoccupation with the disconfirming case (Weed, 2007):

Grounded theory approaches also emphasise the importance of searching for negative or disconfirming cases to challenge emergent analyses . . . and this is also incorporated into the meta-study approach. . . . Similarly, as the iterative process of meta-interpretation develops, theoretical sampling seeks not only to broaden and deepen the analysis, but also to challenge it through seeking alternative points of view and perspectives (p. 19).

Development of a wider choice of methodologies for qualitative evidence synthesis has been accompanied by increasing recognition of the value of identifying the disconfirming case. This stems, at least in part, from a need to demonstrate review findings that go beyond the obvious and that are nuanced to particular contingencies, settings or population subgroups. For example the increasingly popular technique of realist synthesis attempts to explore differences between the outcomes of apparently similar programmes. It seeks to achieve this by unpicking the components of each programme and, indeed, actively exploring dissonance in how, and with whom, they have been implemented (Pawson et al., 2004). Wong and colleagues (2010) describe a realist review where “we deliberately sought out disconfirming data - i.e., data that might refute our provisional candidate theories” (p. 3). May (2006) associates the active searching for deviant or disconfirming cases as characteristic of those formative analyses specifically required where a reviewer is assessing and evaluating complex interventions.

Identification of the disconfirming case may also occupy a place within the specific sub-genre of methodological reviews where vote-counting based approaches are manifestly inadequate (Booth, 2001):

This is particularly the case in methodological reviews where a minority approach may have tremendous potential for development but be otherwise overshadowed by the existence of a substantive corpus of work. This is analogous to the importance of investigating sources of heterogeneity in quantitative meta-analyses. (p. 2)

Procedures for identifying the disconfirming case

Role of the Review Team

Probably the most accessible method for encouraging the active exploration of dissonance and the subsequent incorporation of multiple perspectives is in the constitution of the review team.

For example Lloyd Jones (2004) identifies the value of involving a team of researchers throughout the full range of review processes:

Ideally, however, a meta-synthesis should be undertaken by a team of researchers: the application of multiple perspectives . . . may result in additional insights, and thus in a more complete interpretation of the subject of the review. (Lloyd Jones, 2004, p. 277)

Involvement of multiple team members does not necessarily maximize the prospects that the team is able to identify the disconfirming case. First, constitution of a review team is often “self-limiting”. Implicitly such a team represents a group of individuals with whom a principal investigator believes that they could work. The investigator does not typically select a team based on their variability. A review team might thus share common values or beliefs, whether unspoken or articulated. Second, the commissioner of a review might require that a review team represent different disciplines or areas of methodological expertise rather than necessarily reflecting genuine heterogeneity. In such circumstances the constitution of a team is not specifically a mechanism to identify and explore differences in findings. Indeed a research team more typically reflects diversity in the different views each member holds toward certain types of research or study design (Dixon-Woods et al., 2006a).

Different reviewer backgrounds. Jones (2004) is explicit about the importance of involving reviewers from different backgrounds:

The team-building process begins by recruiting participants (two, three or more per team) from varying backgrounds (professionally as well as demographically) to be immersed in the selected literature, at times ‘line by line’ and hypothesise at each new revelation of dialogic material . . . (Jones, 2004, p. 103).

This description does not address how a review team harnesses such backgrounds nor how best to deploy a review team to capitalize on such advantages. An additional issue that requires

exploration relates to syntheses of published reports in different languages. The use of researchers with different first language backgrounds might enhance the richness of subsequent theoretical thinking as well as assisting in identification of cultural and national nuances.

Reflexivity. The above requirement for different reviewer backgrounds is couched primarily in quantitative terms (i.e., “professionally as well as demographically” (Jones, 2004, p. 103)) rather than as the need to engage with different values or belief systems. There is increasing recognition that the reflexivity of the review team is as important within qualitative evidence synthesis as it is in primary qualitative research. However to date there are few examples where the team has, collectively and individually, formally considered how the diverse values and beliefs that they hold might enhance the richness and complexity of the subsequent analysis.

Team dynamics. Team composition also impacts on the “power dynamic” within the review team. This phenomenon has only previously been considered in the context of research teams in general (Gerstl-Pepin & Gunzenhauser, 2002) but is equally critical in review teams. Actual examples include a proposed review where the aspiring lead reviewer held the joint roles of senior staff member, subject expert and experienced reviewer. Where such a principal investigator appears to “hold all the cards” it is difficult to envisage how a single junior co-reviewer would be empowered to identify dissonance and then, more significantly, to articulate such observations to other team members.

If members of a review team are to harness a positive team dynamic they must pay explicit attention to the dynamics and internal “politics” within their team. All members of the team must be empowered to challenge, and actively resist, interpretations that do not fit comfortably with the data. Indeed team members should recognise such “fresh” insights as potential hypotheses. They should formally charge the original reviewer with re-examining the data specifically for evidence to either “prove” or “disprove” the validity of such insights.

Although the systematic review process conventionally incorporate interaction with steering or advisory groups or consultation with stakeholders it is not usually feasible to ask such groups to fulfil a formal role in looking for contradictions between different data sources. Indeed it is rare for members of steering groups, or other stakeholders, to have either the time or the inclination to look through the supporting data from a major review. It therefore seems likely that the use of supplementary expertise and resource from within the review team is the best route to improving the internal validity of the process. Such procedures should be disassociated from political imperatives to establish the credibility of the review or its subsequent external validity.

Qualitative synthesis does not share with quantitative synthesis the same formal requirement for two reviewers to select articles or extract data from included studies as a protection against bias. Conceivably peer review of a limited random sample of decisions and training and/or piloting of methods might meet the supplementary requirements for quality assurance. This could release this same level of reviewer resource to be harnessed more appropriately in enhancing the quality of analysis and interpretation by maximizing identification of the disconfirming case.

Legitimizing a culture of questioning. A review team must also consider the “culture” within which their review takes place. Typically junior members of a review team are those who conduct the data extraction and preliminary analysis, for financial and practical reasons. Such junior team members might not be encouraged to identify “problem data” that does not seem to fit within the broad sweeping findings or recommendations of a review or within the prevailing “received wisdom” of topic experts (Macintyre & Petticrew, 2000). The time constraints inherent in the systematic review process provide little incentive for exploring diversions or cul de sacs from the main review findings. There are few incentives for spending less overtly productive time in the backwaters of the review topic. Review teams need to put into place methods and opportunities that legitimize the identification and exploration of incongruity. Perhaps a review

team could assign a senior member of the team a formal role as a “champion of incongruity” encouraging the junior team members first to raise and then, with the team’s agreement, to explore any contradictory findings and the implications of these for the review product.

Use of Methods for Identification of Studies

A review team typically has relatively little control of factors relating to team composition and how the team operates (as determined by resources, timescales and available expertise). However team members have much more flexibility in choosing methods that are more likely to lead to identification of the disconfirming case. This is particularly true when a team chooses methods for the sampling, identification and subsequent selection of studies for inclusion. Booth (2001) recognizes that

Literature searching for qualitative systematic reviews should exhibit the following characteristics: Identifying major "schools of thought" in a particular area whilst being alert to the identification of variants, minority views and dissenters. It is particularly important to identify negative or disconfirming cases . . . (p. 4)

Although the comprehensive search strategy remains a gold standard for quantitative systematic reviews there is increasing recognition that, for a qualitative evidence synthesis, it is more critical that a search strategy is selected to match the intended purpose of the review. Methods such as critical interpretive synthesis and realist synthesis emphasize identification of specific items of evidence to address an identified need. Other methods too may derive value from the diversity, rather than the comprehensiveness, of their sampling technique. Purposive or theoretical sampling, in a quest to achieve theoretical saturation, may increase the likelihood that reviewers retrieve disconfirming cases. Indeed recent guidance in the Cochrane Handbook (Noyes et al., 2008) acknowledges that:

a more purposive sampling approach, aiming to provide a holistic interpretation of a phenomenon, where the extent of searching is driven by the need to reach theoretical saturation and the identification of the ‘disconfirming case’, may be more appropriate (p. 20.7)

Downe (2008) similarly describes having “adopted the techniques of theoretical saturation and of searching for disconfirming data, borrowed from grounded theory” (p. 6). Greenhalgh & Peacock (2005) operationalized their approach to meta-narrative as having “mapped 13 different research traditions, compared their conceptual and theoretical approaches, and synthesised the empirical evidence” (p. 1064). Clearly such search methods, including extensive citation tracking, recognize the relative importance of identifying similarities and differences between the different traditions, in preference to systematic identification of all available evidence.

Such considerations, exemplified by Barroso & Sandelowski (2003) within primary qualitative studies, move the model of searching away from *a priori* identification of all relevant research to more contingent iterative approaches to searching:

As the information value of cases depends on analytic goals emerging in the course of study, researchers can only anticipate— prior to entering the field of study—the kind of purposeful sampling they will conduct and the sample composition and size they will likely obtain . . . researchers may not have anticipated the selection of “deviant” cases until they reached a point in the data analysis where the inclusion of such contrasting cases became vital to the development and validation of a typology. (p. 388)

Indeed such considerations raise the possibility of including a subsequent searching phase to attempt systematically to identify disconfirming cases from the literature. Such searching might employ sampling strategies that aim to maximize diversity (e.g., literature from disciplines or schools of thought not hitherto included). Indeed introduction of an additional review phase that

requires maximum variability sampling as a specific validation approach, following initial data analysis, might prove a useful addition to many types of evidence synthesis.

Differential Exposure to the Findings

Typically any reviewer faces the challenging task of moving between the minutiae of data extraction and a more “elevated” stance required for the subsequent synthesis and analysis. Few would argue against the inherent advantages of using someone who has immersed themselves in the data for a period of weeks, or even months, for the ensuing analysis. However it is important not to overlook the associated dangers posed by the cognitive biases of such a reviewer (Petticrew & Roberts, 2006). It could prove advantageous to bring in other members of the review team, with differing levels of prior exposure to the findings, specifically to challenge interim findings and, indeed to co-create new findings. Thorne and Paterson (1998) formalized such a procedure in their description of methods for their synthesis on models of chronic illness. After recounting how “all researchers met to identify differences in their analyses and arrive at an agreement about the analysis” (p. 175) they describe two specific procedures:

the research team identified negative or disconfirming cases in the research when the data did not unanimously support the conclusions of the researcher, . . . and the researchers tested rival hypotheses by seeking explanations, other than the initial researcher's original hypothesis, to describe the data . . . ” (p. 175)

Multiple Readings

Scientists have observed that the human brain is likely to identify similarities ahead of differences (Petticrew & Roberts, 2006). A review team might therefore expect that exposure to multiple readings of study data, both directly from the source article and repackaged as data within data extraction forms, could help them to identify themes overlooked from a preliminary

reading. Indeed O'Connell & Downe (2009) seem to recognize that refutational evidence (citing Noblit & Hare, 1988) is most likely to benefit from within team discussion and repeated reading "The emergent themes were discussed extensively and the studies were reread to consider any evidence that could be considered refutational" (p. 594). Of course there is no inherent virtue in simply re-reading the same data sources. A reviewer needs to employ diverse reading strategies to maximise the likelihood of identifying a finding that does not engage with the prevailing theory.

Different Methods of Analysis

Frequently concerns with the validity of a particular interpretation are addressed within primary studies through the use of *triangulation*. Opportunities to do this are more limited within evidence synthesis, where the reviewer is constrained by the limited data present in published reports. Nevertheless this might prove a useful supplementary strategy. For example O'Connell & Downe (2009) report using a different type of data (namely observational data) to explore their synthesis:

Of particular interest here was [Study A]; this, along with the oldest of the studies [Study B] contained observational data. These were particularly explored to disprove the emerging analysis or any prior reflexive assumptions. (p. 594)

In particular, reviewers might juxtapose process and outcome data to examine if factors that study participants believe critical to the success of an intervention are present or absent in studies that have achieved the most significant effects. Conversely reviewers might consider whether any study features, not necessarily identified initially from the qualitative data, are shared across the most effective or least effective studies. "Sibling" studies, our term for associated studies of different types that share a common temporal and geographical "heritage" (e.g., qualitative

studies, process evaluations or economic evaluations conducted alongside randomized controlled trials or other outcome based studies) are particularly valuable in this context. As indicated above the benefits of examining studies for congruence do not lie solely in the identification of divergence or the discrepant case. A formal “gap analysis” might inform additional analysis, possible components for candidate interventions and the commissioning of future research (Grant & Booth, 2009). For example a research team from the EPPI-Centre at the Institute of Education in London examined the most effective interventions for promoting intake of fruit and vegetables. They identified an almost complete absence of “healthy messages” i.e., that “eating fruit or vegetables is good for you” (Thomas et al., 2003).

Analytic procedures. Review progress meetings and interim reports are structurally designed to focus on commonalities rather than exceptions. A structured report form might be used to counter this tendency as a mechanism for actively encouraging identification of the disconfirming case. In the same way as articles in major biomedical journals articulate “What is already known on this topic” and “What this study adds” interim review reports might consider “What we expected to find from this review but have not” (i.e., the gap analysis) and “What we have found that we did not expect” (i.e., the unexpected finding). The review team might use regular team discussions to explore possible explanations for such findings. They might then identify actions needed to explore such deviations. Documentation of such issues, and the subsequent discussions, would contribute to another important attribute for a qualitative systematic review, namely reflexivity by the review team (Dixon-Woods et al., 2006a; Weed 2007). Such reflexivity is considerably enhanced where team members discuss and document a priori beliefs about the topic under review.

A review team should also consider the need to separate formally the synthesis and analysis stages of the review process. Frequently the review team focuses on bringing studies

together into a new synthesized product. Subsequently the review team needs to test the robustness of the synthesis. However such a stage may fall victim to time pressures. We could draw an analogy with meta-analysis where production of a forest plot does not represent the endpoint for the process. Rather a meta-analyst uses a variety of techniques such as funnel plots, sensitivity analyses and subgroup analyses to examine the robustness of the synthetic product (Egger et al., 1997). Do analogous processes exist within qualitative evidence synthesis?

Qualitative sensitivity and subgroup analyses

Attention to analysis provides an opportunity to identify exceptions to the rule. It also allows a reviewer to explore and to explain such exceptions. A review team should identify circumstances where findings are only present in particular subgroups (cp., subgroup analysis). They should also identify whether any particular study has had a disproportionate influence on the themes present in the final synthesis (cp., sensitivity analysis) (Mills et al., 2005). Would a model or framework still be complete if the team was to remove such a study and its findings from the synthesis? Downe (2007; 2008) suggests that such testing might be formative (i.e., iterative and ongoing), through a constant comparison approach:

We have also adopted the techniques of theoretical saturation and of searching for disconfirming data, borrowed from grounded theory. As we analysed each additional study, we consciously checked if the findings extended or refuted the emerging line of argument synthesis. (p. 6)

Such an approach is undeniably a valuable way for members of a review to explore, and indeed sensitize themselves to, the data. It does not exclude the need for a more summative taking stock of particular groups of studies or groups of findings on completion of the preliminary synthesis.

Voils et al. (2008) describe one variant of such an approach where the study finding, not the study report, becomes the unit of analysis:

Accordingly, we next turned to vote counting, in which a significance level is set as a cutoff, and then each relationship is placed into one of three categories: positive (confirming), negative (disconfirming), or no relationship. The category with the greatest number is then assumed to provide the best estimate of the relationship. (p. 12)

Such an approach differs from others advanced in this article by attempting primarily to quantify dissonance rather than to explain it.

The methodology of qualitative evidence synthesis does not yet include methods for addressing likely publication bias, analogous to use of the funnel plot in quantitative synthesis. Nevertheless members of a review team should consider whether they have captured a rich and diverse range of study features, preferably identified *a priori*, within their synthesis. In particular they need to consider this if the methodology they have employed utilizes theoretical, opportunistic or purposive sampling as an alternative to the comprehensive searching of formal systematic review. Have they identified studies from all the pre-specified disciplines? Have they excluded coverage of any ages or ethnic groups from their population of included studies? Do the findings only originate from those involved in the delivery of the service or intervention being evaluated? Are low resource and high resource settings represented in the sample of included studies? Few authors of published reviews give adequate attention to such issues or consider the implications for their review findings. Indeed it would be useful if authors of published qualitative evidence syntheses examined such issues in a formal and systematic manner within the Limitations subsection of their published reports. In addition, where a review team has conducted a mixed method review they could seek to corroborate findings across both quantitative and qualitative included studies.

Testing of existing theories and frameworks

Obvious dangers are present where the purpose of a qualitative systematic review is to validate existing theory. A reviewer might be encouraged, by temperament, team culture or design of the documentation, to privilege squeezing of extracted data into an existing format or framework in preference to creating new dimensions or concepts. This approach may be particularly attractive given that any new concept could, in theory, necessitate re-coding of previously coded studies if not the actual revisiting of the source article. Under such circumstances reviewers might find it valuable to extract data against two or more competing theories or to compare theory-based with atheoretical interpretations of the data. At a technical level reviewers might find it more useful to extract entire articles to qualitative data analysis software to facilitate independent coding of the same data, including *de novo* coding (literally original coding from scratch) for emerging models or theories, in preference to using a stable set of selective data extracts. Paterson and colleagues (1999) describe how hypothesis testing was essential to developing and refining theory whereby the analytic process:

included developing hypotheses as the fieldwork progressed and testing these hypothetical relationships by means of further data collection and analysis in the search for confirming or disconfirming evidence to support or negate the emerging theory. (p. 791)

Dixon-Woods and colleagues (2006b) describe a similar process but emphasize that this presents challenges to subsequent reporting:

much of the later sampling was directed at testing and purposively challenging the theory as we began to develop it. Again, such forms of searching and sampling do not lend themselves easily to reproducibility or indeed auditability” (p. 11).

The same authors conclude that:

Testing whether the interpretations change in response to different findings will be an important focus for future research, which will also need to evaluate whether apparently disconfirming evidence is the result of methodological flaws or poses a genuine challenge to theory (p. 11).

Similar procedures are well prescribed in concept analysis, a longstanding area of qualitative literature review aimed at the specific generation of theory. Walker & Avant (2005) describe identifying a typical case and a disconfirming case - exactly as a reviewer might seek to do within the context of interpretative qualitative evidence synthesis. Such methods could conceivably form the basis for a methodology for moving from concepts toward theory in a qualitative synthesis.

Methods for Presentation of Findings

This article focuses on identifying disconfirming findings during synthesis and analysis.

Nevertheless similar considerations extend to the shaping process for the final synthetic product.

For example some qualitative evidence syntheses provide a clear auditable process linking findings to their originating studies. Formative presentation of findings among members of the research team might provide an opportunity to assess the extent to which individual studies contribute to the synthesis, whether themes are present in multiple studies, whether particular findings are contradictory or whether particular studies are outliers. Such a situation is described by Sandelowski and colleagues (2007):

we then arranged these abstracted findings to show their topical similarity and thematic diversity, and then referenced each finding with the report(s) from which it was derived. Arranging findings in this way has the advantage of revealing findings that are not there that might theoretically or logically have been expected (p. 107).

Although the authors readily acknowledge that they had not encountered a clear instance of this in their findings they do cite as an example:

the finding that difficulty accepting HIV was a deterrent to adherence . . . This finding suggests an opposite: that acceptance of HIV would favor adherence.” (p. 107)

Subsequently the team reflected on the significance of this process:

Had we not actually found the one report with this finding, we could not have assumed it even though it might make sense to do so. Although theoretically possible findings may be derived from actual empirical findings, they do not constitute actual findings in a metasummary. (Sandelowski et al., 2007, p. 107)

Mechanisms for identifying the disconfirming case must not simply be based on examination of what has been found. There are several reasons why it is helpful to introduce some external frame of reference. First, many synthesis methods claim that “the result is greater than the sum of its parts” (Noblit & Hare, 1988, p. 28). Such a stance maintains that bringing findings together might result in the generation of new constructs. Because such new constructs might be either overarching (i.e., aggregative) or manufactured (i.e., synthetic) they might map imperfectly to the originating studies. Indeed if a synthesis simply represents an aggregation or amalgamation of all its component themes it might yield little added value (or, at the very least, return on investment) beyond representing a map of the literature. We should neither expect all themes to contribute equally to the final synthesis nor any model or framework to explain adequately all identified themes. We should therefore seek to identify positively (i.e., not simply through omission) any themes that are not adequately explained by an emerging model or framework. Such a “best fit” approach contrasts with the more exacting demands of qualitative researchers such as Kidder (1981) who state that negative case analysis strives, through

progressive interpretations, to produce theory that explains all cases without exception. However a best fit approach is accommodated by the pragmatism of those such as Lincoln & Guba (1985) who acknowledge that the process of thematic analysis always holds apparent exceptions.

A second point relates to the fact that, unlike informants in a primary study, findings from published research articles are not produced independently of one another (Dixon-Woods et al., 2006a). Indeed if a research team has followed good practice and conducted a good quality literature search at the beginning of their primary study, it is almost impossible to achieve such independence. Perversely academic incentives, in terms of the receipt of funding, likelihood of publication and individual prestige, depend on identifying and reporting new findings or, at least, in providing a new perspective on existing findings. The influence of findings from other authors is typically constrained in a primary research report to the Introduction or Discussion sections of an article, and is correspondingly less visible in the Findings. An even more worrying form of publication bias is where an author attempts to improve the alleged originality of their study report by failing to acknowledge its precursors at all. Because of these various influences on publication a reviewer or reader must not assume that how a topic is represented in its published literature bears any relation to the true state of the topic. Academic incentives for reporting novelty should encourage an author to isolate and identify a disconfirming case in their individual primary study (as something new about which to report). An attendant danger is that an author might place disproportionate emphasis on findings not previously reported (cp., false positives).

Stakeholder or respondent validation

Another possible, and yet equally controversial, mechanism for identification of a disconfirming case when the findings of a synthesis are presented is stakeholder or respondent validation (Jensen and Allen, 1996). Principal concerns relate to the notion of the synthetic construct. If a

stakeholder or respondent only recognizes a fragment of their truth in the perspective afforded by a meta-synthesis is it still their prerogative to challenge its findings? Should a reviewer privilege a respondent's individual interpretation of a phenomenon in preference to distilled findings from many such individual accounts? Critics of respondent validation within the context of synthesis focus on the difficulty of identifying an appropriate "respondent". Should the respondent perhaps be the individual researcher who has conducted one or more of the included primary research studies? If so, will such a researcher recognize the "truth" of a synthetic construct if they did not identify it when conducting their original data analysis? Respondent validation is only useful if reviewers purposefully solicited authors for data from their studies that, those authors believe, specifically challenges a review's conclusions, rather than asking them to agree with or to endorse the findings from that review. Reviewers could then incorporate seeking of the disconfirming case within such formal consultation. Ultimately the synthesis team should retain control over analysis and interpretation of such supplementary data. At the same time reviewers must acknowledge that commissioners often want to be reassured that the findings from such a synthesis reflect the accurate perceptions of those working within a given topic area. Some form of respondent testing, rather than validation might, in fact, be a useful exercise even if it simply serves to anticipate potential objections to the findings or recommendations of the published evidence synthesis.

Findings from exemplar methods for identifying the disconfirming case

A brief survey of the literature illustrates many potential mechanisms to explore the disconfirming case. However few of these mechanisms have received formal evaluation. The next section presents three reviews that opportunistically explored combinations of the above strategies together with observations from such case studies.

The disconfirming case in a study of protocol based care.

This qualitative review examining the development and implementation of protocol-based care was conducted within a larger project examining the contribution of nurses, midwives and health visitors to such protocols (Patterson et al., 2008; Ilott et al., 2010). Two reviewers, separately and independently, extracted data for a common set of identified articles. One reviewer had become familiar with guidance presenting an idealized 12-step development process for producing protocols (NHS Modernisation Agency, 2003). Using this framework for coding allowed identification of the extent to which data from different studies corresponds with the official guidance. Another reviewer used a grounded theory based approach without reference to the published NHS Modernisation Agency guidance. Both reviewers independently extracted data from the included studies with the former coding deductively using categories derived from the framework and the latter working inductively, using free-form categories assigned on the basis of data extracts. The review team used Joanna Briggs Institute's QARI software (version 2) for organising and compiling data extracts. The team harnessed this software, typically used for aggregative syntheses, in a more interpretive way, providing a ready audit trail for progress from findings to themes and from themes to constructs.

An extensive data report with findings from both reviewers was shared with a visiting scholar who was given several days to digest the findings prior to a review meeting. The two reviewers and the visiting scholar then met to bring together three complementary perspectives. The first perspective derived from the data-driven framework synthesis, the second from the data-driven grounded theory approach whereas the third represented a conceptual, holistic viewpoint. In this way, within the limited constraints of the review, the team maximized the possibility of identifying the disconfirming case.

As expected prominent themes were identified independently by all three reviewers. Reviewer One provided validation for the overarching framework produced by the government-approved model (an idealized account of how the process ought to be). Examination of the published accounts against this framework revealed an absence of unprompted mentions of patient participation in the protocol development process. In the absence of the published framework Reviewer Two had not observed the absence of patient involvement when performing the synthesis. The team considered that patient participation in protocol based care represented government-backed rhetoric rather than the realities of practice.

In addition the published framework portrayed protocol based care as a sequential process. By using a grounded theory approach separated from the published framework Reviewer Two was freed from the “shackles” of such a linear process. The individual steps or components were validated by this independent approach but there was significant variation in the sequencing of the process of protocol development. The second reviewer was also able to identify a much more fluid and iterative approach to the development process. Descriptions in the included reports differed in when, and indeed why, an evidence gathering process took place. Some authors used the literature search process as the starting point for production of their protocol, following prioritization of the topic. Others used the same process as a *post hoc* justification for decisions that they were planning to implement anyway. Single dimensional coding against the published framework as performed by Reviewer One would have missed this important temporal variation. In contrast, in the absence of a frame of reference, Reviewer Two had to identify both the steps of the process and their likely sequence. Furthermore Reviewer Two identified important contextual variables that might prove to be drivers for the protocols. A few protocols were instigated by top down initiatives, most represented a response to a local problem, whereas some could be interpreted as local adaptations of an existing evidence base. The team felt that

these contextual differences might moderate the perceived success of each initiative. Such an observation had been masked from Reviewer One who had been directed towards the mechanics of the development process at the expense of the context.

Reviewer Three brought an additional perspective to the process. Stepping back from the detail of the study reports she was able to detect the presence of significant publication bias. She observed that all the included articles were individual case studies. As such they represented success stories, not evaluations, mostly described by those who were instrumental in implementation of the protocol itself. Such authors had few, if any, incentives for candour or admission of failure, either in acknowledging limited success in reports written for local consumption or in publishing “warts and all” accounts in the peer-reviewed literature. The above account demonstrates that these complementary perspectives do not necessarily contribute to the data synthesis stage but may improve the subsequent analysis by allowing unexpected insights to emerge.

The disconfirming case in a study of attitudes to chemoprevention.

A review team undertook a qualitative evidence synthesis of attitudes toward the taking of potential agents for chemoprevention of colorectal cancer as part of a larger review and economic evaluation (Cooper et al, 2010). Two reviewers extracted data and then a third reviewer was required to look independently at the interpretations of the two reviewers. The team convened a meeting at which members could discuss the findings and offer alternative interpretations stimulated by use of an innovative “best fit” model (Carroll et al, 2011) as a framework for analysis. Unlike the previous case study, this framework was not purpose-specific but it was similar enough to the review topic to yield comparable insights (Huffman, 2002). The team identified inherent contradictions between the external framework and the chemoprevention data (i.e., the external framework was derived from a context involving young women, whereas the

chemoprevention review included both sexes and older age groups). When examining data extracted from included studies against this frame of reference the review team asked: Is this concept likely to transfer to the context of our review? Might this concept be explained away by population, gender or intervention differences from the source study? Although the themes from the existing framework provided a useful starting point, the team could expand on, add to or reduce themes to reflect the chemoprevention data. Such a comparison offered rich opportunities to identify specific additional characteristics present in the older target population.

A second technique involved partial re-analysis of the data against an additional framework identified and suggested by Reviewer Three. The third reviewer examined data from included studies and identified a temporal dimension not captured by the original framework. The team hypothesized that people's attitudes to chemopreventive agents would differ according to whether they were receiving general population advice (precontemplation), were targeted with condition-specific information (contemplation) or were already taking the agents (action). The team reanalysed the data according to this well-recognized transtheoretical model of health behaviour change (Prochaska & Velicer, 1997). This model helped to make sense of otherwise conflicting findings from some of the studies. The fact that important temporal dimensions were initially omitted from both this and the preceding case study suggests that review teams might find it useful to include a prompt such as "Are significant temporal dimensions missing from the model?" whenever they try to validate any logic or process-based model.

Whether such procedures are successful or not depends heavily on the background and conceptual resources of the third reviewer. Such insights only carry the status of hypothesis generation. An additional hypothesis that attitudes might differ between those taking vitamins (associated with being "natural") and chemical substances such as aspirin (associated with being artificial despite its natural origins) received only limited support from the extracted data

(Connor, 2004; Huber et al., 2004). In this case there was not enough evidence either to confirm or refute this tentative hypothesis. Nevertheless this independent observation yielded a useful recommendation for exploration and research. The above demonstrates that a review team might find useful lines of inquiry in existing frameworks and models, even when such models imperfectly represent the topic area under consideration.

The disconfirming case in a study of the student experience of e-learning.

A third review, conducted on a modest budget for the Higher Education Academy (Booth et al, 2009; Carroll et al 2009a, 2009b), nevertheless managed to manufacture opportunities for identification of dissonance. In this case multiple reviewers extracted the data for included studies, reporting the student experience of work-based e-learning, and then a single reviewer completed a preliminary thematic synthesis of data. Other members of the review team, experienced in delivery of e-learning reviewed the data extraction forms and, through a process of memo-ing, made independent observations on the data. The review team considered these notes at a single meeting alongside a preliminary taxonomy of themes. The team discussed the choice of terminology for the themes and sub-themes and the validity of the relationships between sub-themes. Subsequently they undertook minor reorganization of the taxonomy and its associated themes.

In this instance the team felt that the synthesis had not captured the full richness and complexity of the extracted data. A second reviewer revisited the data with a view to identifying potential contradictions or inconsistencies from the original taxonomy. Meanwhile the team validated the initial taxonomy by using it as the basis for an interview schedule with those who deliver work-based e-learning. Observations from these two approaches helped to shape a more sophisticated understanding of the relationship between themes, moving from a taxonomy toward an embryonic model. To cite one such example, work-based e-learners stated a preference for

working through materials at their own pace, so that they could adapt their learning schedule to their individual work and personal circumstances. At the same time such e-learners expressed a desire for group interaction and regular feedback. Additional analysis of these themes revealed inbuilt tensions. It is clearly challenging to design a course where individuals work at their own pace and yet interact effectively as a group. It is equally challenging to offer flexibility for students to work at their own pace and yet to plan a programme of regular assessment and feedback. Again the review team achieved a more nuanced interpretation of potentially conflicting findings having manufactured opportunities to revisit the data and extend the analysis.

Developing a Disconfirming Case Identification Guide.

The three case studies described above, together with the supporting survey of the literature, provide an initial basis for a guide to facilitate and document decisions related to identification of the disconfirming case. Such a toolkit or checklist, from which to select appropriately according to the scope and purpose of the review, the characteristics of the topic and its associated literature and the resources (time, funding and so forth) available to the review team, might therefore include:

- **Strategies relating to the review team:** role of the review team, different reviewer backgrounds, reflexivity, team dynamics and legitimizing a culture of questioning.
- **Strategies relating to the review methods:** use of methods for identification of studies, differential exposure to the findings, multiple readings, different methods of analysis, analytic procedures, qualitative sensitivity and subgroup analyses
- **Strategies relating to use of theory:** Testing of existing theories/frameworks
- **Strategies relating to presentation of findings:** Methods for presentation of findings, stakeholder or respondent validation

Discussion

The approaches described in the three case studies remain tentative and require formal evaluation. Few reviewers report or evaluate techniques that they currently use when exploring contradictions within their data. In these three reviews we employed pragmatic and opportunistic approaches, constrained by available resources and time limitations. The above consideration simply records observations on the processes involved. We have not been able to establish formally the impact of such methods on the overall quality and rigour of the review. Nevertheless it is possible to identify a few common requirements. First it is not sufficient to expect contradictory findings simply to emerge from conventional review processes. The structures and processes of systematic reviews implicitly target consensus and, at least where meta-analysis is concerned, move in the direction of an averaging effect. Instead a review team involved in qualitative evidence synthesis should seek to manufacture a dialectic (Walsh & Downe, 2005). Unlike quantitative reviews, where the role of the dual observer is primarily to counter bias, a second observer in qualitative synthesis could be employed to pursue persistently contradiction and disconfirmation. Such a quest requires that the review team use reflexivity more extensively and explicitly (as in the method of critical interpretive synthesis) acknowledging that findings from reviews are essentially subjective and open to interpretation (Dixon-Woods et al., 2006b):

CIS demands constant reflexivity on the part of authors of reviews. Authors are charged with making conscientious and thorough searches, with making fair and appropriate selections of materials, with seeking disconfirming evidence and other challenges to the emergent theory, and with ensuring that the theory they generate is, while critically informed, plausible given the available evidence. (p. 9)

A review team might use numerous and diverse methods to create such a dialectic. Where resources permit they might utilize additional reviewers at judicious stages of analysis and interpretation. This is likely to prove particularly fruitful by selecting reviewers with varying levels of prior familiarity with the topic area and different personal inventories of theoretical and practical approaches to synthesis, analysis and interpretation. A review team might use literature searches to operationalize maximum variation sampling by accessing disciplines or schools of thought that emphasize diversity and dissonance. Reviewers might juxtapose theoretical and atheoretical approaches to synthesis and compare and contrast their findings. They might use external models and frameworks to structure the initial data extraction or to provide an alternative overlay against which to reinterpret and reconcile contradictory findings. Above all review teams need to establish a culture where contradiction and discrepancy are legitimized and, indeed, actively encouraged.

Although the potential for such approaches is more limited where only a single reviewer is involved, for example where a doctoral student is undertaking a synthesis, many of the above techniques are still applicable. A supervisor or mentor may explicitly adopt a role that seeks to identify and explore dissonance, as embodied in the stance of a “devil’s advocate”. Opportunities to present interim findings to wider groups with diverse backgrounds can also be engineered to be more interactive and less affirmative. Optimally such dissonance should be handled as a property of the diversity of the literature and not as an inadequacy of the analysis or analyst.

Such a "toolkit" for handling contradictory data is most likely to be appropriate to meta-synthesis approaches with a realist orientation (e.g., thematic synthesis) that are intended to generate outcomes that are concrete and definitive. For example, it is likely to be appropriate to use some or most of these procedures where the intention is to generate a list of key dimensions of the phenomenon under study. In those syntheses that feature a more idealist and constructivist

orientation the notion of contradictory data is not relevant. Such approaches, including meta-narrative and critical interpretive synthesis (CIS), are conceptual in focus, with the essence of the overall picture being what is considered significant. Nevertheless, even in such cases, the review team should engineer the presence of procedures to identify other types of contradiction and rival explanations. Under such circumstances it is perhaps more accurately to describe this as “the dissenting voice” as opposed to “the disconfirming case”. The research community needs to explore the potential usefulness of some of the individual procedures itemized above for these specific circumstances.

Conclusion

A brief consideration of these issues helps the reader to recognize how many processes codified in conventional systematic review methodology implicitly steer towards consensus. Typically two reviewers, with a shared understanding of the task in hand, review titles or abstracts or complete data extraction with a view to reaching such a consensus. Where such a consensus is not readily apparent the team might refer queries to a third person, acting as arbiter, charged with agreeing a final common position on the specific case or issue. To increase confidence in the process a review team might undertake inter-rater reliability (kappa) tests. The final account for the report of the systematic review thus represents a shared narrative to which all contributors are expected to subscribe. The outcome of such consensual methods contrasts with what might result if two reviewers were required to construct *separate* narratives from the same data sources and then to bring the two narratives together and place them in juxtaposition.

Within qualitative systematic review methodology there is increasing recognition that reviewers choose the type of synthesis based primarily on whether an established and largely accepted framework exists for the literature under consideration (Noyes et al., 2008). The

emerging orthodoxy states that where concepts are “secure” it might be most appropriate for reviewers to use a framework or model based approach, for example using framework synthesis as a variant of framework analysis (Dixon-Woods et al., 2006a, 2006b). In contrast, where concepts are still undergoing development and refinement more interpretive approaches, such as those employed by meta-ethnography or grounded theory, are likely to be more appropriate (Dixon-Woods et al., 2006a, 2006b). Our experience suggests an additional consideration when a review team is selecting the appropriate “methodological mix” – that is the need for reviewers to explore actively any diversity within the data. Certain scenarios, such as the need to challenge an official position or a theory that is seen to monopolize a particular field, might require reviewers to undertake separate framework and grounded theory based analyses and then to compare the findings from each resultant synthesis. Similarly independent use of two competing frameworks or models might result in a richer degree of interpretation, revealing perhaps the inadequacy of a particular model. Such procedures might even contribute to an optimal fusion of two or more models within a “meta-model” (Carroll et al., 2011).

In a multi-author reflection on the methodological agenda Noblit (Thorne et al, 2004) observes that:

Refutation syntheses seemed to me to hold the most promise, because they invited us to consider that ideas were, indeed, contested via ethnographies, seemingly indisputable to me. Moreover, they invite considerable creative and critical talents to the work of synthesis, yet this form of synthesis is rare, to my knowledge (p. 1349).

However while other authors similarly affirm the benefits of seeking the disconfirming case few propose mechanisms by which a review team might achieve this in practice. We hope that the above discussion leads to more systematic use and evaluation of strategies for identifying dissonance within evidence synthesis through consideration of the constitution of the review

team, use of contrasting methods for analysis, use of theoretical models as a catalyst for exploration, and clear and auditable methods for presentation of data. In continuing this quest many proposed solutions are likely to lie, not within the emerging discipline of qualitative evidence synthesis but rather, on the well-trod foothills of primary qualitative data analysis.

References

- Barroso, J., & Sandelowski, M. (2003). Sample reporting in qualitative studies of women with HIV infection. *Field Methods*, 15(4), 386-404. doi: 10.1177/1525822X03257392
- Booth, A. (2001). Cochrane or cock-eyed? how should we conduct systematic reviews of qualitative research? *Paper presented at the Qualitative Evidence-based Practice Conference, Taking a Critical Stance*. Coventry University, May 14-16.
<http://www.leeds.ac.uk/educol/documents/00001724.htm>
- Booth, A., Carroll, C., Papaioannou, D., Sutton, A., & Wong, R. (2009). Applying findings from a systematic review of workplace-based e-learning: implications for health information professionals. *Health Information & Libraries Journal*, 26(1), 4-21. doi: 10.1111/j.1471-1842.2008.00834.x
- Carroll, C., Booth, A., Papaioannou, D., Sutton, A., & Wong, R. (2009a). Enhancing the experience of e-learning among working students: a systematic review with thematic analysis. *Impact: Journal of Applied Research in Workplace E-learning*, 1(1), 80-96.
- Carroll, C., Booth, A., Papaioannou, D., Sutton, A., & Wong, R. (2009b). UK health-care professionals' experience of on-line learning techniques: A systematic review of qualitative data. *Journal of Continuing Education in the Health Professions*, 29(4), 235-241. doi: 10.1002/chp.20041
- Carroll C, Booth A, Cooper K. (2011). A worked example of "best fit" framework synthesis: A systematic review of views concerning the taking of some potential chemopreventive agents. *BMC Medical Research Methodology*, 11, 29. doi: 10.1186/1471-2288-11-29
- Connor, L.H. (2004). Relief, risk and renewal: mixed therapy regimens in an Australian suburb. *Social Science & Medicine*, 59, 1695–1705. doi: 10.1016/j.socscimed.2004.01.030
- Cooper, K., Squires, H., Carroll, C., Papaioannou, D., Booth, A., Logan, R. F., Maguire C, Hind D & Tappenden, P. (2010). Chemoprevention of colorectal cancer: systematic review and

- economic evaluation. *Health Technology Assessment*, 14(32), 1-206. doi: 10.3310/hta14320
- Dixon-Woods, M., Bonas, S., Booth, A., Jones, D.R., Miller, T., Shaw, R.L., Smith, J., Sutton, A., & Young, B. (2006a). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, 6, 27-44. doi: 10.1177/1468794106058867
- Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R. & Sutton, A.J. (2006) Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups *BMC Medical Research Methodology*, 6, 35. doi: 10.1186/1471-2288-6-35
- Dong, N., Maynard, R., & Perez-Johnson, I. (2008) Averaging effect sizes within and across studies of interventions aimed at improving child outcomes. *Child Development Perspectives*, 2(3), 187-197. doi: 10.1111/j.1750-8606.2008.00064.x
- Downe, S., Simpson, L., & Trafford, K. (2007). Expert intrapartum maternity care: a metasynthesis. *Journal of Advanced Nursing*, 57(2), 127-140. doi: 10.1111/j.1365-2648.2006.04079.x
- Downe, S. (2008). Metasynthesis: a guide to knitting smoke. *Evidence Based Midwifery*, 6(1), 4-8. http://www.doctoralmidwiferysociety.org/Portals/c8d3e3f8-9c01-4bf5-abd9-3fd6b4c510ae/Downe_S.pdf
- Egger, M., Smith, G.D., & Phillips, A.N. (1997). Meta-analysis: principles and procedures. *BMJ*, 315(7121), 1533-1537. doi: 10.1136/bmj.315.7121.1533
- Gerstl-Pepin, C. & Gunzenhauser, M. (2002). Collaborative team ethnography and the paradoxes of interpretation. *Qualitative Studies in Education*, 15 (2), 137-154. doi: 10.1080/09518390110111884
- Grant, M.J., & Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91-108. doi: 10.1111/j.1471-1842.2009.00848.x
- Greenhalgh, T. & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*, 331, 1064-1065. doi: 10.1136/bmj.38636.593461.68

- Huber, R., Koch, D., Beiser, I., Zschocke, I., & Luedtke, R. (2004). Experience and attitudes towards CAM – a survey of internal and psychosomatic patients in a German university hospital. *Alternative Therapies in Health & Medicine*, 10, 32–36.
- Huffman, S. (2002). *Can marketing of multiple vitamin/mineral supplements reach the poor?* The Vitaldía Project, Bolivia. http://pdf.usaid.gov/pdf_docs/PNACCS889.pdf
- Ilott, I., Booth, A., Rick, J., & Patterson, M. (2010). (2010) How do nurses, midwives and health visitors contribute to protocol-based care? A synthesis of the UK literature. *International Journal of Nursing Studies*, 47(6), 770-780. doi: 10.1016/j.ijnurstu.2009.12.023
- Jensen, L.A. & Allen, M.N. (1996). Meta-synthesis of qualitative findings. *Qualitative Health Research*, 6, 553-560. doi: 10.1177/104973239600600407
- Jones, K. (2004). Mission drift in qualitative research, or moving toward a systematic review of qualitative studies, moving back to a more systematic narrative review. *The Qualitative Report*, 9 (1), 95-112. <http://www.nova.edu/ssss/QR/QR9-1/jones.pdf>
- Kidder, L.H. (1981). Qualitative research and quasi-experimental frameworks. In *Scientific inquiry and the social sciences*, edited by M. B. Brewer & B. E. Collins, 226–256. San Francisco: Jossey-Bass.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills: Sage.
- Lloyd Jones, M. (2004) Application of systematic review methods to qualitative research: practical issues. *Journal of Advanced Nursing*, 48(3), 271–278. doi: 10.1111/j.1365-2648.2004.03196.x
- Macintyre, S., & Petticrew, M. (2000) Good intentions and received wisdom are not enough. *Journal of Epidemiology & Community Health*, 54(11), 802-803. doi:10.1136/jech.54.11.802
- May, C. (2006). A rational model for assessing and evaluating complex interventions in health care. *BMC Health Services Research*, 6, 86. doi:10.1186/1472-6963-6-86
- Mills, E., Jadad, A.R., Ross, C., & Wilson, K. (2005). Systematic review of qualitative studies exploring parental beliefs and attitudes toward childhood vaccination identifies common barriers to vaccination. *Journal of Clinical Epidemiology*, 58(11), 1081-1088. doi:10.1016/j.jclinepi.2005.09.002
- NHS Modernisation Agency/National Institute for Clinical Excellence. (2003) *Protocol-based care: underpinning improvement*. London: NSHMA/NICE.

http://www.institute.nhs.uk/quality_and_service_improvement_tools/quality_and_service_improvement_tools/protocol_based_care.html

- Noblit, G.W., & Hare, R.D. (1988). *Meta-ethnography: synthesising qualitative studies* Sage: Thousand Oaks, CA. 88pp
- Noyes, J., Popay, J., Pearson, A., Hannes, K., & Booth, A. (2008) Chapter 20: Qualitative research and Cochrane reviews. In: Higgins, J.P.T., & Green, S. (editors), *Cochrane handbook for systematic reviews of interventions*, version 5.0.1 (updated September 2008). The Cochrane Collaboration. Available from www.cochrane-handbook.org.
- O'Connell, R., & Downe, S. (2009). A metasynthesis of midwives' experience of hospital practice in publicly funded settings: compliance, resistance and authenticity. *Health* (London), 13(6), 589-609. doi: 10.1177/1363459308341439
- Papaioannou, D., Sutton, A., Carroll, C., Booth, A., & Wong, R. (2010) Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information & Libraries Journal*, 27(2), 114-22. doi: 10.1111/j.1471-1842.2009.00863.x
- Paterson, B.L., Thorne, S., & Dewis, M. (1998). Adapting to and managing diabetes. *Image Journal of Nursing Scholarship*, 30(1), 57-62. doi: 10.1111/j.1547-5069.1998.tb01237.x
- Paterson, B., Thorne, S., Crawford, J., & Tarko, M. (1999). Living with diabetes as a transformational experience. *Qualitative Health Research*, 9(6), 786-802. doi: 10.1177/104973299129122289
- Paterson, B.L., Thorne, S.E., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research. a practical guide to meta-analysis and meta-synthesis*. Thousand Oaks, CA: Sage Publications.
- Patterson, M., Rick, J., Ilott, I., Lekka, C., Dixon, S., & Oluboyede, Y. (2008). The contribution of nurses, midwives and health visitors to protocol-based care and its variants, and the impact of their contribution on patient and staff outcomes, quality and costs of care. Report for the National Institute for Health Research Service Delivery and Organisation programme, June 2008. <http://www.sdo.nihr.ac.uk/files/project/79-final-report.pdf>
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. 2d ed. Newbury Park, CA: Sage.

- Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2004). *Realist synthesis: an introduction*. Manchester: ESRC Research Methods Programme. Available at:
<http://www.ccsr.ac.uk/methods/publications/documents/RMPmethods2.pdf>
- Petticrew, M., & Roberts, H. (2006). How to appraise the studies: an introduction to assessing study quality. In: *Systematic reviews in the social sciences: a practical guide*. Oxford: Blackwell; 2006.p. 130
- Prochaska, J., & Velicer, W. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, 12, 38–48. doi: 10.4278/0890-1171-12.1.38
- Sandelowski, M., Barroso, J., & Voils, C.I. (2007). Using qualitative metasummary to synthesize qualitative and quantitative descriptive findings. *Research in Nursing & Health*, 30(1), 99-111. doi: 10.1002/nur.20176
- Thomas, J., Sutcliffe, K., Harden, A., Oakley, A., Oliver, S., Rees, R., Brunton G, & Kavanagh J. (2003). *Children and healthy eating: A systematic review of barriers and facilitators*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
http://eppi.ioe.ac.uk/EPPIWebContent/hp/reports/healthy_eating02/Final_report_web.pdf
- Thorne, S., & Paterson, B. (1998). Shifting images of chronic illness. *Image Journal of Nursing Scholarship*, 30 (2), 173–178. doi: 10.1111/j.1547-5069.1998.tb01275.x
- Thorne, S., Jensen, L., Kearney, M.H., Noblit, G., & Sandelowski, M. (2004). Qualitative metasynthesis: reflections on methodological orientation and ideological agenda. *Qualitative Health Research*, 14 (10), 1342-65. doi: 10.1177/1049732304269888
- Voils, C.I., Sandelowski, M., Barroso, J. & Hasselblad, V. (2008). Making sense of qualitative and quantitative findings in mixed research synthesis studies. *Field Methods*, 20, 3-25. doi: 10.1177/1525822X07307463
- Walker, L.O. & Avant, K.C. (2005) *Strategies for theory construction in nursing*, 4th edn. Pearson Prentice Hall, Upper Saddle River, NJ.
- Walsh, D., & Downe, S. (2005). Meta-synthesis method for qualitative research: a literature review. *Journal of Advanced Nursing*, 50(2), 204-211. doi: 10.1111/j.1365-2648.2005.03380.x

- Weed, M. (2007). A potential method for the interpretive synthesis of qualitative research: issues in the development of 'meta-interpretation'. *International Journal of Social Research Methodology*, 11(1), 13-28. doi: 10.1080/13645570701401222
- Wong, G., Greenhalgh, T., & Pawson, R. (2010). Internet-based medical education: a realist review of what works, for whom and in what circumstances. *BMC Medical Education*, 10, 12. doi: 10.1186/1472-6920-10-12